

B1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number  
**WO 02/04490 A2**

(51) International Patent Classification<sup>7</sup>: C07K 14/00

(21) International Application Number: PCT/US01/21432

(22) International Filing Date: 6 July 2001 (06.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

|            |                            |    |
|------------|----------------------------|----|
| 60/216,801 | 7 July 2000 (07.07.2000)   | US |
| 60/216,803 | 7 July 2000 (07.07.2000)   | US |
| 60/218,233 | 14 July 2000 (14.07.2000)  | US |
| 60/220,046 | 21 July 2000 (21.07.2000)  | US |
| 60/220,739 | 26 July 2000 (26.07.2000)  | US |
| 60/222,824 | 4 August 2000 (04.08.2000) | US |

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **DAS, Debopriya** [IN/US]; 1267 Parkington Avenue, Sunnyvale, CA 94087 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue #1, Redwood City, CA 94025 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **KHAN, Farrah, A.** [IN/US]; 3617 Central Road, # 102, Glenview, IL 60025 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **YAO, Monique,**

**G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **EL-LIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **NGUYEN, Dannie, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **BAUGN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(74) Agent: **HAMLET-COX, Diana**; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/04490 A2

(54) Title: LIPID METABOLISM MOLECULES

(57) Abstract: The invention provides human lipid metabolism molecules (LMM) and polynucleotides which identify and encode LMM. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of LMM.

## LIPID METABOLISM MOLECULES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of lipid metabolism molecules and to the use of these sequences in the diagnosis, treatment, and prevention of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, skin disorders, and cardiovascular disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of lipid metabolism molecules.

### BACKGROUND OF THE INVENTION

Lipids are water-insoluble, oily or greasy substances that are soluble in nonpolar solvents such as chloroform or ether. Neutral fats (triacylglycerols) serve as major fuels and energy stores. Polar lipids, such as phospholipids, sphingolipids, glycolipids, and cholesterol, are key structural components of cell membranes. (Lipid metabolism is reviewed in Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY; Lehninger, A. (1982) Principles of Biochemistry, Worth Publishers, Inc. New York NY; and ExPASy "Biochemical Pathways" index of Boehringer Mannheim World Wide Web site, "<http://www.expasy.ch/cgi-bin/search-biochem-index>".)

Fatty acids are long-chain organic acids with a single carboxyl group and a long non-polar hydrocarbon tail. Long-chain fatty acids are essential components of glycolipids, phospholipids, and cholesterol, which are building blocks for biological membranes, and of triglycerides, which are the major energy store in animals. Long-chain fatty acids are also substrates for eicosanoid production. Eicosanoids, including prostaglandins, prostacyclin, thromboxanes, and leukotrienes, are 20-carbon signaling molecules with roles in pain, fever, and inflammation. The precursor of all eicosanoids is arachidonate, which is derived from phospholipids and from diacylglycerols. Long-chain fatty acids are also important in the functional modification of certain complex carbohydrates and proteins. 16-carbon and 18-carbon fatty acids are the most common.

Fatty acid synthesis, or lipogenesis, is similar in prokaryotes and eukaryotes. In prokaryotes, seven enzymes catalyze the various synthetic steps. In the first step, acetyl-CoA carboxylase (ACC) synthesizes malonyl-CoA from acetyl-CoA and bicarbonate. Subsequently, malonyl transacylase attaches malonyl-CoA to the 4'-phosphopantetheine prosthetic group of acyl carrier protein (ACP), producing malonyl-ACP. ACPs serve as scaffolds which are responsible for holding the components of the growing fatty acid chain in close proximity to the appropriate enzyme. Acetyl-ACP is produced by the action of acetyl transacylase on acetyl-CoA and ACP. Acetyl-ACP and malonyl-ACP are condensed and reduced by the enzyme  $\beta$ -ketoacyl-ACP synthase, forming D-3-hydroxybutyryl-ACP. D-3-hydroxybutyryl-ACP then undergoes further enzymatic reactions,

including dehydration and a second reduction, to produce butyryl-ACP, the end product of the first round of elongation. Subsequent rounds of elongation proceed until a 16-carbon chain is produced. This 16-carbon chain is cleaved from ACP to produce palmitate.

- Eukaryotic lipogenesis involves the same biochemical reactions as prokaryotic lipogenesis.
- 5 ACC again catalyzes the first step in this process, producing malonyl-CoA. However, the enzymes which catalyze the remaining reactions are covalently linked into a single polypeptide chain, referred to as the multifunctional enzyme fatty acid synthase (FAS). FAS catalyzes the synthesis of palmitate from acetyl-CoA and malonyl-CoA (Wakil, S.J., et al. (1989) *Biochem* 28:4523-4530). Further elongation, as well as unsaturation, of palmitate by accessory enzymes of the endoplasmic reticulum
- 10 produces the variety of long chain fatty acids required by the individual cell (Siggaard-Andersen, M., et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11027-11031). Prior to their incorporation into other cellular lipids, fatty acids are activated by long-chain fatty acid-CoA ligase. This enzyme contains an AMP-binding domain signature (Etchegaray, A. (1998) *Biochem. Mol. Biol. Int* 44:235-243).

- Lipogenesis occurs at a low and fairly constant level in the cells of most mammalian tissues.
- 15 Carbohydrate intake in excess of that required for immediate energy needs is stored as triacylglycerol. When carbohydrate intake is less than required for immediate energy needs, stored triacylglycerol is hydrolyzed and utilized as fuel. Fat stored in liver and adipose triglycerides may be released by hydrolysis and transported in the blood. Free fatty acids are transported in the blood by albumin. Triacylglycerols and cholesterol esters in the blood are transported in lipoprotein particles. The
- 20 particles consist of a core of hydrophobic lipids surrounded by a shell of polar lipids and apolipoproteins. The protein components serve in the solubilization of hydrophobic lipids and also contain cell-targeting signals. Lipoproteins include chylomicrons, chylomicron remnants, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). There is a strong inverse correlation between the levels
- 25 of plasma HDL and risk of premature coronary heart disease.

- Degradation of saturated and unsaturated fatty acids begins with the activation of the fatty acid by long-chain fatty acid-CoA ligase. Degradation occurs via mitochondrial and peroxisomal beta-oxidation enzymes, which sequentially remove two-carbon units from CoA-activated fatty acids. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids while the
- 30 auxiliary pathway performs additional steps required for the degradation of unsaturated fatty acids. The pathways of mitochondrial and peroxisomal beta-oxidation use similar enzymes, but have different substrate specificities and functions. Mitochondria oxidize short-, medium-, and long-chain fatty acids to produce energy for cells, particularly in cardiac and skeletal muscle. Peroxisomes oxidize medium-, long-, and very-long-chain fatty acids, dicarboxylic fatty acids, branched fatty
- 35 acids, prostaglandins, xenobiotics, and bile acid intermediates to facilitate the excretion of toxic

lipophilic carboxylic acids and to prepare very-long-chain fatty acids for mitochondrial beta-oxidation. Enzymes involved in beta-oxidation include acyl CoA synthetase/long chain fatty acid - CoA ligase, carnitine acyltransferase, acyl CoA dehydrogenases, enoyl CoA hydratases, L-3-hydroxyacyl CoA dehydrogenase,  $\beta$ -ketothiolase, 2,4-dienoyl CoA reductase, and isomerase.

5        Triacylglycerols, also known as triglycerides and neutral fats, are major energy stores in animals. Triacylglycerols are esters of glycerol with three fatty acid chains. Glycerol-3-phosphate is produced from dihydroxyacetone phosphate by the enzyme glycerol phosphate dehydrogenase or from glycerol by glycerol kinase. Fatty acid-CoA's are produced from fatty acids by fatty acyl-CoA synthetases. Glycerol-3-phosphate is acylated with two fatty acyl-CoA's by the enzyme glycerol  
10        phosphate acyltransferase to give phosphatidate. Phosphatidate phosphatase converts phosphatidate to diacylglycerol, which is subsequently acylated to a triacylglycerol by the enzyme diglyceride acyltransferase. Phosphatidate phosphatase and diglyceride acyltransferase form a triacylglycerol synthetase complex bound to the ER membrane.

      A major class of phospholipids are the phosphoglycerides, which are composed of a glycerol  
15        backbone, two fatty acid chains, and a phosphorylated alcohol. Phosphoglycerides are components of cell membranes. Principal phosphoglycerides are phosphatidyl choline, phosphatidyl ethanolamine, phosphatidyl serine, phosphatidyl inositol, and diphosphatidyl glycerol. Many enzymes involved in phosphoglyceride synthesis are associated with membranes (Meyers, R.A. (1995) Molecular Biology and Biotechnology, VCH Publishers Inc., New York NY, pp. 494-501). Phosphatidate is converted to  
20        CDP-diacylglycerol by the enzyme phosphatidate cytidyltransferase (ExPASy ENZYME EC 2.7.7.41). Transfer of the diacylglycerol group from CDP-diacylglycerol to serine to yield phosphatidyl serine, or to inositol to yield phosphatidyl inositol, is catalyzed by the enzymes CDP-diacylglycerol-serine O-phosphatidyltransferase and CDP-diacylglycerol-inositol 3-  
      phosphatidyltransferase, respectively (ExPASy ENZYME EC 2.7.8.8; ExPASy ENZYME EC  
25        2.7.8.11). The enzyme phosphatidyl serine decarboxylase catalyzes the conversion of phosphatidyl serine to phosphatidyl ethanolamine, using a pyruvate cofactor (Voelker, D.R. (1997) *Biochim. Biophys. Acta* 1348:236-244). Phosphatidyl choline is formed using diet-derived choline by the reaction of CDP-choline with 1,2-diacylglycerol, catalyzed by diacylglycerol  
      cholinephosphotransferase (ExPASy ENZYME 2.7.8.2).

30        Cholesterol, composed of four fused hydrocarbon rings with an alcohol at one end, moderates the fluidity of membranes in which it is incorporated. In addition, cholesterol is used in the synthesis of steroid hormones such as cortisol, progesterone, estrogen, and testosterone. Bile salts derived from cholesterol facilitate the digestion of lipids. Cholesterol in the skin forms a barrier that prevents excess water evaporation from the body. Farnesyl and geranylgeranyl groups, which are derived from  
35        cholesterol biosynthesis intermediates, are post-translationally added to signal transduction proteins



such as Ras and protein-targeting proteins such as Rab. These modifications are important for the activities of these proteins (Guyton, A.C. (1991) Textbook of Medical Physiology, W.B. Saunders Company, Philadelphia PA, pp. 760-763; Stryer, supra, pp. 279-280, 691-702, 934). Mammals obtain cholesterol derived from both de novo biosynthesis and the diet.

5 Sphingolipids are an important class of membrane lipids that contain sphingosine, a long chain amino alcohol. They are composed of one long-chain fatty acid, one polar head alcohol, and sphingosine or a sphingosine derivative such as ceramide. Ceramide is formed by acylation of sphingosine by sphingosine acetyltransferase. The three classes of sphingolipids are sphingomyelins, cerebrosides, and gangliosides. Sphingomyelin is formed by the conversion of ceramide and  
10 phosphatidyl choline by ceramide choline phosphotransferase. Cerebrosides are synthesized by the linkage of glucose or galactose to ceramide by a transferase. Sequential addition of sugar residues to ceramide by transferase enzymes yields gangliosides. Sphingomyelins are abundant in the myelin sheath surrounding nerve cells, while galactocerebrosides are characteristic of the brain. Other cerebrosides are found in non-neuronal tissues, and gangliosides are abundant in the brain, but are  
15 also found in non-neuronal tissues.

A number of sphingolipid metabolites, including ceramide and sphingosine, have been implicated as lipid second messenger molecules that are critically involved in such cell functions as apoptosis, proliferation, differentiation, and inflammation (reviewed by Liu, G. et al. (1999) Crit. Rev. Lab. Sci. 36:511-573). Many of these second messengers derive from the lysosomal degradation  
20 of glycosphingolipids. Glycosphingolipids are hydrolyzed by an exohydrolase in combination with a sphingolipid activator protein (SAP, or saposin), a non-enzymatic glycoprotein cofactor. SAPs are synthesized as a single precursor which is cleaved to give four homologous polypeptides (SAP-A, SAP-B, SAP-C, and SAP-D). SAPs seem to function by direct activation of the enzyme, or by lifting the sphingolipid substrate out of the plane of the membrane. A SAP-C deficiency causes a variant  
25 form of Gaucher's disease (Harzer, K. et al. (1989) Eur. J. Pediatr. 149:31-39).

Eicosanoids, including prostaglandins, prostacyclin, thromboxanes, and leukotrienes, are 20-carbon molecules derived from fatty acids. Eicosanoids are signaling molecules which have roles in pain, fever, and inflammation. The precursor of all eicosanoids is arachidonate, which is generated from phospholipids by phospholipase A<sub>2</sub> and from diacylglycerols by diacylglycerol lipase.  
30 Leukotrienes are produced from arachidonate by the action of lipoxygenases.

Within cells, fatty acids are transported by cytoplasmic fatty acid binding proteins (Online Mendelian Inheritance in Man (OMIM) \*134650 Fatty Acid-Binding Protein 1, Liver; FABP1). Diazepam binding inhibitor (DBI), also known as endozepine and acyl CoA-binding protein, is an endogenous  $\gamma$ -aminobutyric acid (GABA) receptor ligand which is thought to down-regulate the  
35 effects of GABA. DBI binds medium- and long-chain acyl-CoA esters with very high affinity and

may function as an intracellular carrier of acyl-CoA esters (OMIM \*125950 Diazepam Binding Inhibitor; DBI; PROSITE PDOC00686 Acyl-CoA-binding protein signature).

Fat stored in liver and adipose triglycerides may be released by hydrolysis and transported in the blood. Free fatty acids are transported in the blood by albumin. Triacylglycerols and cholesterol esters in the blood are transported in lipoprotein particles. The particles consist of a core of hydrophobic lipids surrounded by a shell of polar lipids and apolipoproteins. The protein components serve in the solubilization of hydrophobic lipids and also contain cell-targeting signals. Lipoproteins include chylomicrons, chylomicron remnants, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). There is a strong inverse correlation between the levels of plasma HDL and risk of premature coronary heart disease.

Mitochondrial and peroxisomal beta-oxidation enzymes degrade saturated and unsaturated fatty acids by sequential removal of two-carbon units from CoA-activated fatty acids. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids while the auxiliary pathway performs additional steps required for the degradation of unsaturated fatty acids. The pathways of mitochondrial and peroxisomal beta-oxidation use similar enzymes, but have different substrate specificities and functions. Mitochondria oxidize short-, medium-, and long-chain fatty acids to produce energy for cells. Mitochondrial beta-oxidation is a major energy source for cardiac and skeletal muscle. In liver, it provides ketone bodies to the peripheral circulation when glucose levels are low as in starvation, endurance exercise, and diabetes (Eaton, S. et al. (1996) *Biochem. J.* 320:345-357). Peroxisomes oxidize medium-, long-, and very-long-chain fatty acids, dicarboxylic fatty acids, branched fatty acids, prostaglandins, xenobiotics, and bile acid intermediates. The chief roles of peroxisomal beta-oxidation are to shorten toxic lipophilic carboxylic acids to facilitate their excretion and to shorten very-long-chain fatty acids prior to mitochondrial beta-oxidation (Mannaerts, G.P. and P.P. Van Veldhoven (1993) *Biochimie* 75:147-158). Enzymes involved in beta-oxidation include acyl CoA synthetase, carnitine acyltransferase, acyl CoA dehydrogenases, enoyl CoA hydratases, L-3-hydroxyacyl CoA dehydrogenase,  $\beta$ -ketothiolase, 2,4-dienoyl CoA reductase, and isomerase.

Three classes of lipid metabolism molecules are discussed in further detail. The three classes are lipases, phospholipases and lipoxigenases.

#### Lipases

Triglycerides are hydrolyzed to fatty acids and glycerol by lipases. Adipocytes contain lipases that break down stored triacylglycerols, releasing fatty acids for export to other tissues where they are required as fuel. Lipases are widely distributed in animals, plants, and prokaryotes. Triglyceride lipases (ExPASy ENZYME EC 3.1.1.3), also known as triacylglycerol lipases and

tributyrase, hydrolyze the ester bond of triglycerides. In higher vertebrates there are at least three tissue-specific isozymes including gastric, hepatic, and pancreatic lipases. These three types of lipases are structurally closely related to each other as well as to lipoprotein lipase. The most conserved region in gastric, hepatic, and pancreatic lipases is centered around a serine residue which is also present in lipases of prokaryotic origin. Mutation in the serine residue renders the enzymes inactive. Gastric, hepatic, and pancreatic lipases hydrolyze lipoprotein triglycerides and phospholipids. Gastric lipases in the intestine aid in the digestion and absorption of dietary fats. Hepatic lipases are bound to and act at the endothelial surfaces of hepatic tissues. Hepatic lipases also play a major role in the regulation of plasma lipids. Pancreatic lipase requires a small protein cofactor, colipase, for efficient dietary lipid hydrolysis. Colipase binds to the C-terminal, non-catalytic domain of lipase, thereby stabilizing an active conformation and considerably increasing the overall hydrophobic binding site. Deficiencies of these enzymes have been identified in man, and all are associated with pathologic levels of circulating lipoprotein particles (Gargouri, Y. et al. (1989) *Biochim. Biophys. Acta* 1006:255-271; Connelly, P.W. (1999) *Clin. Chim. Acta* 286:243-255; van Tilbeurgh, H. et al. (1999) *Biochim Biophys Acta* 1441:173-184).

Lipoprotein lipases (ExPASy ENZYME EC 3.1.1.34), also known as clearing factor lipases, diglyceride lipases, or diacylglycerol lipases, hydrolyze triglycerides and phospholipids present in circulating plasma lipoproteins, including chylomicrons, very low and intermediate density lipoproteins, and high-density lipoproteins (HDL). Together with pancreatic and hepatic lipases, lipoprotein lipases (LPL) share a high degree of primary sequence homology. Both lipoprotein lipases and hepatic lipases are anchored to the capillary endothelium via glycosaminoglycans and can be released by intravenous administration of heparin. LPLs are primarily synthesized by adipocytes, muscle cells, and macrophages. Catalytic activities of LPLs are activated by apolipoprotein C-II and are inhibited by high ionic strength conditions such as 1 M NaCl. LPL deficiencies in humans contribute to metabolic diseases such as hypertriglyceridemia, HDL2 deficiency, and obesity (Jackson, R.L. (1983) in *The Enzymes* (Boyer, P.D., ed.) Vol. XVI, pp. 141-186, Academic Press, New York NY; Eckel, R.H. (1989) *New Engl. J. Med.* 320:1060-1068).

#### Phospholipases

Phospholipases, a group of enzymes that catalyze the hydrolysis of membrane phospholipids, are classified according to the bond cleaved in a phospholipid. They are classified into PLA1, PLA2, PLB, PLC, and PLD families. Phospholipases are involved in many inflammatory reactions by making arachidonate available for eicosanoid biosynthesis. More specifically, arachidonic acid is processed into bioactive lipid mediators of inflammation such as lyso-platelet-activating factor and eicosanoids. The synthesis of arachidonic acid from membrane phospholipids is the rate-limiting step in the biosynthesis of the four major classes of eicosanoids (prostaglandins, prostacyclins,

thromboxanes and leukotrienes) which are involved in pain, fever, and inflammation (Kaiser, E. et al. (1990) Clin. Biochem. 23:349-370). Furthermore, leukotriene-B<sub>4</sub> is known to function in a feedback loop which further increases PLA<sub>2</sub> activity (Wijkander, J. et al. (1995) J. Biol. Chem. 270:26543-26549).

5       The secretory phospholipase A<sub>2</sub> (PLA<sub>2</sub>) superfamily comprises a number of heterogeneous enzymes whose common feature is to hydrolyze the sn-2 fatty acid acyl ester bond of phosphoglycerides. Hydrolysis of the glycerophospholipids releases free fatty acids and lysophospholipids. PLA<sub>2</sub> activity generates precursors for the biosynthesis of biologically active lipids, hydroxy fatty acids, and platelet-activating factor. PLA<sub>2</sub>s were first described as components  
10 of snake venoms, and were later characterized in numerous species. PLA<sub>2</sub>s have traditionally been classified into several major groups and subgroups based on their amino acid sequences, divalent cation requirements, and location of disulfide bonds. The PLA<sub>2</sub>s of Groups I, II, and III consist of low molecular weight, secreted, Ca<sup>2+</sup>-dependent proteins. Group IV PLA<sub>2</sub>s are primarily 85-kDa, Ca<sup>2+</sup>-dependent cytosolic phospholipases. Finally, a number of Ca<sup>2+</sup>-independent PLA<sub>2</sub>s have been  
15 described, which comprise Group V (Davidson, F.F. and E.A. Dennis (1990) J. Mol. Evol. 31:228-238; and Dennis, E.F. (1994) J. Biol Chem. 269:13057-13060).

The first PLA<sub>2</sub>s to be extensively characterized were the Group I, II, and III PLA<sub>2</sub>s found in snake and bee venoms. These venom PLA<sub>2</sub>s share many features with mammalian PLA<sub>2</sub>s including a common catalytic mechanism, the same Ca<sup>2+</sup> requirement, and conserved primary and tertiary  
20 structures. In addition to their role in the digestion of prey, the venom PLA<sub>2</sub>s display neurotoxic, myotoxic, anticoagulant, and proinflammatory effects in mammalian tissues. This diversity of pathophysiological effects is due to the presence of specific, high affinity receptors for these enzymes on various cells and tissues (Lambeau, G. et al. (1995) J. Biol. Chem. 270:5534-5540).

PLA<sub>2</sub>s from Groups I, IIA, IIC, and V have been described in mammalian and avian cells,  
25 and were originally characterized by tissue distribution, although the distinction is no longer absolute. Thus, Group I PLA<sub>2</sub>s were found in the pancreas, Group IIA and IIC were derived from inflammation-associated tissues (e.g., the synovium), and Group V were from cardiac tissue. The pancreatic PLA<sub>2</sub>s function in the digestion of dietary lipids and have been proposed to play a role in cell proliferation, smooth muscle contraction, and acute lung injury. The Group II inflammatory  
30 PLA<sub>2</sub>s are potent mediators of inflammatory processes and are highly expressed in serum and synovial fluids of patients with inflammatory disorders. These Group II PLA<sub>2</sub>s are found in most human cell types assayed and are expressed in diverse pathological processes such as septic shock, intestinal cancers, rheumatoid arthritis, and epidermal hyperplasia. A Group V PLA<sub>2</sub> has been cloned from brain tissue and is strongly expressed in heart tissue. A human PLA<sub>2</sub> was recently  
35 cloned from fetal lung, and based on its structural properties, appears to be the first member of a new

group of mammalian PLA2s, referred to as Group X. Other PLA2s have been cloned from various human tissues and cell lines, suggesting a large diversity of PLA2s (Chen, J. et al. (1994) *J. Biol. Chem.* 269:2365-2368; Kennedy, B.P. et al. (1995) *J. Biol. Chem.* 270: 22378-22385; Komada, M. et al. (1990) *Biochem. Biophys. Res. Commun.* 168:1059-1065; Cupillard, L. et al. (1997) *J. Biol. Chem.* 272:15745-15752; and Nalefski, E.A. et al. (1994) *J. Biol. Chem.* 269:18239-18249).

Phospholipases B (PLB) (ExPASy ENZYME EC 3.1.1.5), also known as lysophospholipase, lecithinase B, or lysolecithinase are widely distributed enzymes that metabolize intracellular lipids, and occur in numerous isoforms. Small isoforms, approximately 15-30 kD, function as hydrolases; large isoforms, those exceeding 60 kD, function both as hydrolases and transacylases. A particular substrate for PLBs, lysophosphatidylcholine, causes lysis of cell membranes when it is formed or imported into a cell. PLBs are regulated by lipid factors including acylcarnitine, arachidonic acid, and phosphatidic acid. These lipid factors are signaling molecules important in numerous pathways, including the inflammatory response (Anderson, R. et al. (1994) *Toxicol. Appl. Pharmacol.* 125:176-183; Selle, H. et al. (1993); *Eur. J. Biochem.* 212:411-416).

Phospholipase C (PLC) (ExPASy ENZYME EC 3.1.4.10) plays an important role in transmembrane signal transduction. Many extracellular signaling molecules including hormones, growth factors, neurotransmitters, and immunoglobulins bind to their respective cell surface receptors and activate PLCs. The role of an activated PLC is to catalyze the hydrolysis of phosphatidyl-inositol-4, 5-bisphosphate (PIP<sub>2</sub>), a minor component of the plasma membrane, to produce diacylglycerol and inositol 1, 4, 5-trisphosphate (IP<sub>3</sub>). In their respective biochemical pathways, IP<sub>3</sub> and diacylglycerol serve as second messengers and trigger a series of intracellular responses. IP<sub>3</sub> induces the release of Ca<sup>2+</sup> from internal cellular storage, and diacylglycerol activates protein kinase C (PKC). Both pathways are part of transmembrane signal transduction mechanisms which regulate cellular processes which include secretion, neural activity, metabolism, and proliferation.

Several distinct isoforms of PLC have been identified and are categorized as PLC-beta, PLC-gamma, and PLC-delta. Subtypes are designated by adding Arabic numbers after the Greek letters, eg. PLC- $\beta$ -1. PLCs have a molecular mass of 62-68 kDa, and their amino acid sequences show two regions of significant similarity. The first region designated X has about 170 amino acids, and the second or Y region contains about 260 amino acids.

The catalytic activities of the three isoforms of PLC are dependent upon Ca<sup>2+</sup>. It has been suggested that the binding sites for Ca<sup>2+</sup> in the PLCs are located in the Y-region, one of two conserved regions. The hydrolysis of common inositol-containing phospholipids, such as phosphatidylinositol (PI), phosphatidylinositol 4-monophosphate (PIP), and phosphatidylinositol 4, 5-bisphosphate (PIP<sub>2</sub>), by any of the isoforms yields cyclic and noncyclic inositol phosphates (Rhee,

S.G. and Y.S. Bae (1997) *J. Biol. Chem.* 272:15045-15048).

All mammalian PLCs contain a pleckstrin homology (PH) domain which is about 100 amino acids in length and is composed of two antiparallel beta sheets flanked by an amphipathic alpha helix. PH domains target PLCs to the membrane surface by interacting with either the beta/gamma subunits of G proteins or PIP2 (PROSITE PDOC50003).

Phospholipase D (PLD) (ExPASy ENZYME EC 3.1.4.4), also known as lecithinase D, lipophosphodiesterase II, and choline phosphatase catalyzes the hydrolysis of phosphatidylcholine and other phospholipids to generate phosphatidic acid. PLD plays an important role in membrane vesicle trafficking, cytoskeletal dynamics, and transmembrane signal transduction. In addition, the activation of PLD is involved in cell differentiation and growth (reviewed in Liscovitch, M. (2000) *Biochem. J.* 345:401-415).

PLD is activated in mammalian cells in response to diverse stimuli that include hormones, neurotransmitters, growth factors, cytokines, activators of protein kinase C, and agonists binding to G-protein-coupled receptors. At least two forms of mammalian PLD, PLD1 and PLD2, have been identified. PLD1 is activated by protein kinase C alpha and by the small GTPases ARF and RhoA. (Houle, M.G. and S. Bourgoin (1999) *Biochim. Biophys. Acta* 1439:135-149). PLD2 can be selectively activated by unsaturated fatty acids such as oleate (Kim, J.H. (1999) *FEBS Lett.* 454:42-46).

#### Lipoxygenases

Lipoxygenases (ExPASy ENZYME EC 1.13.11.12) are non-heme iron-containing enzymes that catalyze the dioxygenation of certain polyunsaturated fatty acids such as lipoproteins. Lipoxygenases are found widely in plants, fungi, and animals. Several different lipoxygenase enzymes are known, each having a characteristic oxidation action. In animals, there are specific lipoxygenases that catalyze the dioxygenation of arachidonic acid at the carbon-3, 5, 8, 11, 12, and 15 positions. These enzymes are named after the position of arachidonic acid that they dioxygenate. Lipoxygenases have a single polypeptide chain with a molecular mass of ~75-80 kDa in animals. The proteins have an N-terminal-barrel domain and a larger catalytic domain containing a single atom of non-heme iron. Oxidation of the ferric enzyme to an active form is required for catalysis (Yamamoto, S. (1992) *Biochim. Biophys. Acta* 1128:117-131; Brash, A.R. (1999) *J. Biol. Chem.* 274:23679-23682). A variety of lipoxygenase inhibitors exist and are classified into five major categories according to their mechanism of inhibition. These include antioxidants, iron chelators, substrate analogues, lipoxygenase-activating protein inhibitors, and, finally, epidermal growth factor-receptor inhibitors.

3-Lipoxygenase, also known as e-LOX-3 or Aloxe3 has recently been cloned from murine epidermis. Aloxe3 resides on mouse chromosome 11, and the deduced amino acid sequence for

Aloxe3 is 54% identical to the 12-lipoxygenase sequences (Kinzig, A. (1999) Genomics 58:158-164).

5-Lipoxygenase (5-LOX, ExPASy ENZYME EC 1.13.11.34), also known as arachidonate:oxygen 5-oxidoreductase, is found primarily in white blood cells, macrophages, and mast cells. 5-LOX converts arachidonic acid first to 5-hydroperoxyeicosatetraenoic acid (5-HPETE) and then to leukotriene (LTA4 (5,6-oxido-7,9,11,14-eicosatetraenoic acid)). Subsequent conversion of leukotriene A4 by leukotriene A4 hydrolase yields the potent neutrophil chemoattractant leukotriene B4. Alternatively, conjugation of LTA4 with glutathione by leukotriene C4 synthase plus downstream metabolism leads to the cysteinyl leukotrienes that influence airway reactivity and mucus secretion, especially in asthmatics. Most lipoxygenases require no other cofactors or proteins for activity. In contrast, the mammalian 5-LOX requires calcium and ATP, and is activated in the presence of a 5-LOX activating protein (FLAP). FLAP itself binds to arachidonic acid and supplies 5-LOX with substrate (Lewis, R.A. et al. (1990) New Engl. J. Med. 323:645-655). The expression levels of 5-LOX and FLAP are found to be increased in the lungs of patients with plexogenic (primary) pulmonary hypertension (Wright, L. et al. (1998) Am. J. Respir. Crit. Care Med. 157:219-229).

12-Lipoxygenase (12-LOX, ExPASy ENZYME: EC 1.13.11.31) oxygenates arachidonic acid to form 12-hydroperoxyeicosatetraenoic acid (12-HPETE). Mammalian 12-lipoxygenases are named after the prototypical tissues of their occurrence (hence, the leukocyte, platelet, or epidermal types). Platelet-type 12-LOX has been found to be the predominant isoform in epidermal skin specimens and epidermoid cells. Leukocyte 12-LOX was first characterized extensively from porcine leukocytes and was found to have a rather broad distribution in mammalian tissues by immunochemical assays. Besides tissue distribution, the leukocyte 12-LOX is distinguished from the platelet-type enzyme by its ability to form 15-HPETE, in addition to 12-HPETE from arachidonic acid substrate. Leukocyte 12-LOX is highly related to 15-lipoxygenase (15-LOX) in that both are dual specificity lipoxygenases, and they are about 85% identical in primary structure in higher mammals. Leukocyte 12-LOX is

Lipid metabolism is involved in human diseases and disorders. Increased synthesis of long-chain fatty acids occurs in neoplasms including those of the breast, prostate, ovary, colon and endometrium. There is a strong inverse correlation between the levels of plasma HDL and risk of premature coronary heart disease. In the arterial disease atherosclerosis, fatty lesions form on the inside of the arterial wall. These lesions promote the loss of arterial flexibility and the formation of blood clots (Guyton, A.C. Textbook of Medical Physiology (1991) W.B. Saunders Company, Philadelphia PA pp. 760-763). Steatosis, or fatty liver, is characterized by the accumulation of triglycerides in the liver and may occur in association with a variety of conditions including alcoholism, diabetes, obesity, and prolonged parenteral nutrition. Steatosis may lead to fibrosis and cirrhosis of the liver. Eicosanoids, including prostaglandins, prostacyclins, thromboxanes, and leukotrienes, are important mediators of inflammatory responses. Cell lines derived from human ovarian, endometrial, breast, colorectal, and prostatic cancers show increased fatty acid synthesis and a preference for use of endogenously-synthesized fatty acids over dietary lipids as fuel for cellular function (Pizer, E.S., et al. (1996) *Cancer Res* 56:1189-1193). In a rat model for non-insulin-dependent diabetes mellitus, increased hepatic FAS activity results in hypertriglyceridemia (Kazumi, T., et al. (1997), *Endocr J* 44:239-245). High-fat low-carbohydrate diets regulate the expression of FAS, and inhibit FAS activity. Loss of this regulation is associated with weight gain and the development of obesity (Hillgartner, F.B., et al. (1995) *Physiol Rev.* 75:47-76).

Sphingolipids are also associated with various disease states. In Tay-Sachs disease, the GM<sub>2</sub> ganglioside (a sphingolipid) accumulates in lysosomes of the central nervous system due to a lack of the enzyme N-acetylhexosaminidase. Patients suffer nervous system degeneration leading to early death (Fauci, A.S. et al. (1998) Harrison's Principles of Internal Medicine McGraw-Hill, New York NY p. 2171). Inhibition of glycosphingolipid biosynthesis has been shown to improve outcomes in an animal model of Tay-Sachs disease (reviewed in Koter, T. and K. Sandhoff (1998) *Brain Pathol.* 8:79-100). The Niemann-Pick diseases are caused by defects in lipid metabolism. Niemann-Pick diseases types A and B are caused by accumulation of sphingomyelin (a sphingolipid) and other lipids in the central nervous system due to a defect in the enzyme sphingomyelinase, leading to neurodegeneration and lung disease. Niemann-Pick disease type C results from a defect in cholesterol transport, leading to the accumulation of sphingomyelin and cholesterol in lysosomes and a secondary reduction in sphingomyelinase activity. Neurological symptoms such as grand mal seizures, ataxia, and loss of previously learned speech, manifest 1-2 years after birth. A mutation in the NPC protein, which contains a putative cholesterol-sensing domain, was found in a mouse model of Niemann-Pick disease type C (Fauci, supra, p. 2175; Loftus, S.K. et al. (1997) *Science* 277:232-235).

PLAs are implicated in a variety of disease processes. For example, PLAs are found in the



pancreas, in cardiac tissue, and in inflammation-associated tissues. Pancreatic PLAs function in the digestion of dietary lipids and have been proposed to play a role in cell proliferation, smooth muscle contraction, and acute lung injury. Inflammatory PLAs are potent mediators of inflammatory processes and are highly expressed in serum and synovial fluids of patients with inflammatory disorders. Additionally, inflammatory PLAs are found in most human cell types and are expressed in diverse pathological processes such as septic shock, intestinal cancers, rheumatoid arthritis, and epidermal hyperplasia.

The role of PLBs in human tissues has been investigated in various research studies. Hydrolysis of lysophosphatidylcholine by PLBs causes lysis in erythrocyte membranes (Selle, *supra*). Similarly, Endresen, M.J. et al. (1993; Scand. J. Clin. Invest. 53:733-739) reported that the increased hydrolysis of lysophosphatidylcholine by PLB in pre-eclamptic women causes release of free fatty acids into the sera. In renal studies, PLB was shown to protect  $\text{Na}^+, \text{K}^+$ -ATPase from the cytotoxic and cytolytic effects of cyclosporin A (Anderson, *supra*).

Lipases, phospholipases, and lipoxygenases are thought to contribute to complex diseases, such as atherosclerosis, obesity, arthritis, asthma, and cancer, as well as to single gene defects, such as Wolman's disease and Type I hyperlipoproteinemia.

The discovery of new lipid metabolism molecules, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, skin disorders, disorders of lipid metabolism, and cardiovascular disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of lipid metabolism molecules.

### SUMMARY OF THE INVENTION

The invention features purified polypeptides, lipid metabolism molecules, referred to collectively as "LMM" and individually as "LMM-1," "LMM-2," "LMM-3," "LMM-4," "LMM-5," "LMM-6," "LMM-7," and "LMM-8." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-8.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-8. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:9-16.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid

sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional LMM, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional LMM, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional LMM, comprising

administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:9-16, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a

polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing  
5 particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a  
10 reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be  
15 used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## 20 DEFINITIONS

"LMM" refers to the amino acid sequences of substantially purified LMM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of  
25 LMM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LMM either by directly interacting with LMM or by acting on components of the biological pathway in which LMM participates.

An "allelic variant" is an alternative form of the gene encoding LMM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in  
30 polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

35 "Altered" nucleic acid sequences encoding LMM include those sequences with deletions,

insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as LMM or a polypeptide with at least one functional characteristic of LMM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding LMM, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding LMM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent LMM. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of LMM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of LMM. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LMM either by directly interacting with LMM or by acting on components of the biological pathway in which LMM participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind LMM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin,



and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic LMM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding LMM or fragments of LMM may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated

DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

|    | Original Residue | Conservative Substitution |
|----|------------------|---------------------------|
|    | Ala              | Gly, Ser                  |
|    | Arg              | His, Lys                  |
| 15 | Asn              | Asp, Gln, His             |
|    | Asp              | Asn, Glu                  |
|    | Cys              | Ala, Ser                  |
|    | Gln              | Asn, Glu, His             |
|    | Glu              | Asp, Gln, His             |
| 20 | Gly              | Ala                       |
|    | His              | Asn, Arg, Gln, Glu        |
|    | Ile              | Leu, Val                  |
|    | Leu              | Ile, Val                  |
|    | Lys              | Arg, Gln, Glu             |
| 25 | Met              | Leu, Ile                  |
|    | Phe              | His, Met, Leu, Trp, Tyr   |
|    | Ser              | Cys, Thr                  |
|    | Thr              | Ser, Val                  |
|    | Trp              | Phe, Tyr                  |
| 30 | Tyr              | His, Phe, Trp             |
|    | Val              | Ile, Leu, Thr             |

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which

retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

5 A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

10 A "fragment" is a unique portion of LMM or the polynucleotide encoding LMM which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10,  
15 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the  
20 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:9-16 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:9-16, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:9-16 is useful, for  
25 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:9-16 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:9-16 and the region of SEQ ID NO:9-16 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-8 is encoded by a fragment of SEQ ID NO:9-16. A fragment of  
30 SEQ ID NO:1-8 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-8. For example, a fragment of SEQ ID NO:1-8 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-8. The precise length of a fragment of SEQ ID NO:1-8 and the region of SEQ ID NO:1-8 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the  
35 fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

5 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous  
10 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes  
15 in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some  
20 alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e  
25 sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

30 Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

35 *Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

5       Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment  
10   length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

      "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

15       The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

      "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific  
20   hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive  
25   conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

30       Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and  
35   conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al.

(1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of LMM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of LMM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of LMM. For example, modulation

may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of LMM.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an LMM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of LMM.

"Probe" refers to nucleic acid sequences encoding LMM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for



example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs  
5 can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to  
10 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from  
15 megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection  
20 programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both  
25 unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

30 A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have  
35 been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a

recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing LMM, nucleic acids encoding LMM, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

5 A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

15 A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

25 A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an

"allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## 20 THE INVENTION

The invention is based on the discovery of new human lipid metabolism molecules (LMM), the polynucleotides encoding LMM, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, skin disorders, disorders of lipid metabolism, and cardiovascular disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3

shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

5 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the  
10 MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these  
15 properties establish that the claimed polypeptides are lipid metabolism molecules. For example, SEQ ID NO:1 is 63% identical to rat phospholipase C delta-4 (GenBank ID g571466) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $3.6e-204$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains phosphatidylinositol-specific phospholipase X and Y domains, a  
20 C2 domain, and a pleckstrin homology(PH) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:1 is a phospholipase C. SEQ ID NO:3 is 44% identical to human cytosolic phospholipase A2 beta (GenBank ID g3811347) as determined by the Basic Local  
25 Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $4.0e-156$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a lysophospholipase catalytic domain and a C2 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) The data from BLAST and HMM  
30 analyses indicate that SEQ ID NO:3 is a phospholipase A2. SEQ ID NO:4 is 39% identical to mouse schwannoma-associated protein (GenBank ID g2565396), a phospholipase D homolog, as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $2.9e-74$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also contains phospholipase D active site motifs as determined by  
35 searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM

database of conserved protein family domains. (See Table 3.) These data provide corroborative evidence that SEQ ID NO:4 is a phospholipase D. SEQ ID NO:6 is 50% identical to human cytosolic phospholipase A2 beta (GenBank ID g3811347) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $1.5e-205$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a C2 domain and a lysophospholipase catalytic domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provide further corroborative evidence that SEQ ID NO:6 is a phospholipase A2. SEQ ID NO:7 is 38% identical to fadD36, a long-chain fatty acid-CoA ligase homolog from mycobacterium (GenBank ID g1929067) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $8.9e-58$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a putative AMP-binding domain signature as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is a long-chain fatty acid-CoA ligase. SEQ ID NO:2, SEQ ID NO:5, and SEQ ID NO:8 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-8 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:9-16 or that distinguish between SEQ ID NO:9-16 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 3349205H1 is the

identification number of an Incyte cDNA sequence, and BRAITUT24 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70796816V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2110818) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>YYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

| Prefix         | Type of analysis and/or examples of programs   |
|----------------|--|
| GNN, GFG, ENST | Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK). |
| GBI            | Hand-edited analysis of genomic sequences.   |

|    |  |
|----|--|
| FL | Stitched or stretched genomic sequences (see Example V). |
|----|--|

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses LMM variants. A preferred LMM variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the LMM amino acid sequence, and which contains at least one functional or structural characteristic of LMM.

The invention also encompasses polynucleotides which encode LMM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:9-16, which encodes LMM. The polynucleotide sequences of SEQ ID NO:9-16, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding LMM. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding LMM. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:9-16 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:9-16. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of LMM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding LMM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the



polynucleotide sequence of naturally occurring LMM, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode LMM and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring LMM under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding LMM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding LMM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode LMM and LMM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding LMM or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:9-16 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M.

(1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding LMM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer

controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode LMM may be cloned in recombinant DNA molecules that direct expression of LMM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express LMM.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter LMM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of LMM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding LMM may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, LMM itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques.

(See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of LMM, or any part thereof, may be altered during direct synthesis and/or  
5 combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by  
10 sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active LMM, the nucleotide sequences encoding LMM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and  
15 inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding LMM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding LMM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding LMM and its initiation codon and upstream regulatory sequences are inserted  
20 into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of  
25 enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding LMM and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in  
30 vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding LMM. These include, but are not limited to, microorganisms such as bacteria transformed  
35 with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with

yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster  
5 (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses,  
10 adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.)  
15 The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding LMM. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding LMM can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPT1  
20 plasmid (Life Technologies). Ligation of sequences encoding LMM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of LMM are needed, e.g. for the production of  
25 antibodies, vectors which direct high level expression of LMM may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of LMM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH  
30 promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

35 Plant systems may also be used for expression of LMM. Transcription of sequences

encoding LMM may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.)  
5 These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases  
10 where an adenovirus is used as an expression vector, sequences encoding LMM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses LMM in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma  
15 virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino  
20 polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of LMM in cell lines is preferred. For example, sequences encoding LMM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous  
25 expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using  
30 tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk<sup>-</sup>* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic,  
35 or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to

methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *par* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which  
5 alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system.  
10 (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding LMM is inserted within a marker gene sequence, transformed cells containing sequences encoding LMM can be identified by the absence of marker gene function. Alternatively, a  
15 marker gene can be placed in tandem with a sequence encoding LMM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding LMM and that express LMM may be identified by a variety of procedures known to those of skill in the art. These  
20 procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of LMM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques  
25 include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on LMM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and  
30 Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled  
35 hybridization or PCR probes for detecting sequences related to polynucleotides encoding LMM

include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding LMM, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase  
5 such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

10 Host cells transformed with nucleotide sequences encoding LMM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode LMM may be designed to contain signal sequences which  
15 direct secretion of LMM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or  
20 "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

25 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding LMM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric LMM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of LMM activity. Heterologous protein and  
30 peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and  
35 metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity



purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the LMM encoding sequence and the heterologous protein sequence, so that LMM may be cleaved away from the heterologous moiety following purification.

- 5 Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled LMM may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These  
10 systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

LMM of the present invention or fragments thereof may be used to screen for compounds that specifically bind to LMM. At least one and up to a plurality of test compounds may be screened  
15 for specific binding to LMM. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of LMM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):  
20 Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which LMM binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express LMM, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or  
25 E. coli. Cells expressing LMM or cell membrane fractions which contain LMM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either LMM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example,  
30 the assay may comprise the steps of combining at least one test compound with LMM, either in solution or affixed to a solid support, and detecting the binding of LMM to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a  
35 solid support.

LMM of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of LMM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for LMM activity, wherein LMM is combined with at least one test compound, and the activity of LMM in the presence of a test compound is compared with the activity of LMM in the absence of the test compound. A change in the activity of LMM in the presence of the test compound is indicative of a compound that modulates the activity of LMM. Alternatively, a test compound is combined with an in vitro or cell-free system comprising LMM under conditions suitable for LMM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of LMM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding LMM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding LMM may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding LMM can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding LMM is injected into animal ES cells, and the injected sequence

integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress LMM, e.g., by secreting LMM in its milk, may also  
5 serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of LMM and lipid metabolism molecules. In addition, the expression of LMM is closely associated with paraganglionic tumor, skin tissue, fetal brain tissue, brain tissue, prostate  
10 tumor tissue, and fibroblastic and epidermal tissues. Therefore, LMM appears to play a role in cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, skin disorders, disorders of lipid metabolism, and cardiovascular disorders. In the treatment of disorders associated with increased LMM expression or activity, it is desirable to decrease the expression or activity of LMM. In the treatment of disorders associated with decreased LMM expression or  
15 activity, it is desirable to increase the expression or activity of LMM.

Therefore, in one embodiment, LMM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LMM. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in  
20 particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other  
25 extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-  
30 Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other  
35 neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis,

inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal

5 degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-

10 candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's

15 syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal

20 carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis,

25 Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein

30 obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a skin disorder, such as dermatitis, eczema, ichthyosis, keratosis, psoriasis, scleroderma, and skin atrophy; a disorder of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine

35 deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency,

hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipid

5 adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cardiovascular disorder such as

10 arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis,

15 congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary

20 embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis,

25 hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung

30 transplantation.

In another embodiment, a vector capable of expressing LMM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LMM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified LMM in

35 conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent

a disorder associated with decreased expression or activity of LMM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of LMM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LMM including, but not limited to, those listed above.

In a further embodiment, an antagonist of LMM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of LMM. Examples of such disorders include, but are not limited to, those cancer, neurological disorders, autoimmune/inflammatory disorders, gastrointestinal disorders, skin disorders, disorders of lipid metabolism, and cardiovascular disorders described above. In one aspect, an antibody which specifically binds LMM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express LMM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding LMM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of LMM including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of LMM may be produced using methods which are generally known in the art. In particular, purified LMM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind LMM. Antibodies to LMM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with LMM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in

humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to LMM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or  
5 fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of LMM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to LMM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not  
10 limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the  
15 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce LMM-specific single  
20 chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as  
25 disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for LMM may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of  
30 the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either  
35 polyclonal or monoclonal antibodies with established specificities are well known in the art. Such

immunoassays typically involve the measurement of complex formation between LMM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering LMM epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

5 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for LMM. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of LMM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their  
10 affinities for multiple LMM epitopes, represents the average affinity, or avidity, of the antibodies for LMM. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular LMM epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the LMM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations  
15 with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of LMM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

20 The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of LMM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and  
25 guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

In another embodiment of the invention, the polynucleotides encoding LMM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA,  
30 RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding LMM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding LMM. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense  
35 sequences into appropriate target cells can be used. Antisense sequences can be delivered



intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral  
5 vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding LMM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined  
15 immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii)  
20 express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in LMM expression or regulation causes disease, the expression of LMM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

30 In a further embodiment of the invention, diseases or disorders caused by deficiencies in LMM are treated by constructing mammalian expression vectors encoding LMM and introducing these vectors by mechanical means into LMM-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene  
35 transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev.*

Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of LMM include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA),  
5 PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). LMM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA  
10 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous  
15 gene encoding LMM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method  
20 (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to LMM expression are treated by constructing a retrovirus vector consisting of (i) the  
25 polynucleotide encoding LMM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc.  
30 Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R.  
35 et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining

retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding LMM to cells which have one or more genetic abnormalities with respect to the expression of LMM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding LMM to target cells which have one or more genetic abnormalities with respect to the expression of LMM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing LMM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation

of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding LMM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for LMM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of LMM-coding RNAs and the synthesis of high levels of LMM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of LMM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding LMM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for  
5 secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques  
10 for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding LMM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines,  
15 cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs  
20 and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a  
25 compound which is effective in altering expression of a polynucleotide encoding LMM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective  
30 compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased LMM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding LMM may be therapeutically useful, and in the treatment of disorders associated with decreased LMM expression or activity, a compound which specifically promotes expression of the  
35 polynucleotide encoding LMM may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding LMM is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding LMM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding LMM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition

which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may  
5 consist of LMM, antibodies to LMM, and mimetics, agonists, antagonists, or inhibitors of LMM.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

10 Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the  
15 lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination  
20 of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising LMM or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, LMM or a fragment thereof may be joined to a short cationic N-  
25 terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs,  
30 monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example LMM or fragments thereof, antibodies of LMM, and agonists, antagonists or inhibitors of LMM, which  
35 ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by

standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

- 10       The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about  $0.1 \mu\text{g}$  to  $100,000 \mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

- 20       Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

- 25       In another embodiment, antibodies which specifically bind LMM may be used for the diagnosis of disorders characterized by expression of LMM, or in assays to monitor patients being treated with LMM or agonists, antagonists, or inhibitors of LMM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for LMM include methods which utilize the antibody and a label to detect LMM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

- 35       A variety of protocols for measuring LMM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of LMM expression. Normal or standard values for LMM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to LMM under



conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of LMM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

5 In another embodiment of the invention, the polynucleotides encoding LMM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of LMM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of  
10 LMM, and to monitor regulation of LMM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding LMM or closely related molecules may be used to identify nucleic acid sequences which encode LMM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a  
15 conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding LMM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the LMM encoding sequences. The hybridization probes of the subject  
20 invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:9-16 or from genomic sequences including promoters, enhancers, and introns of the LMM gene.

Means for producing specific hybridization probes for DNAs encoding LMM include the cloning of polynucleotide sequences encoding LMM or LMM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may  
25 be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding LMM may be used for the diagnosis of disorders  
30 associated with expression of LMM. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and  
35 uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral

neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema,

5 epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central

10 nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),

15 akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis,

20 autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,

25 hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral,

30 bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis,

35 pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis,

passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a skin disorder, such as dermatitis, eczema, ichthyosis, keratosis, psoriasis, scleroderma, and skin atrophy; a disorder of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases,

pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar  
5 proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation. The polynucleotide sequences encoding LMM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect  
10 altered LMM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding LMM may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding LMM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a  
15 suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding LMM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to  
20 monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of LMM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding LMM, under conditions suitable for hybridization or amplification.  
25 Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

30 Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

35 With respect to cancer, the presence of an abnormal amount of transcript (either under- or

overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding LMM may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding LMM, or a fragment of a polynucleotide complementary to the polynucleotide encoding LMM, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding LMM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding LMM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of LMM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of

interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray  
5 can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the  
10 activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

15 In another embodiment, LMM, fragments of LMM, or antibodies specific for LMM may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of  
20 gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of  
25 transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines,  
30 biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental  
35 compounds. All compounds induce characteristic gene expression patterns, frequently termed

molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical

density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for LMM to quantify the levels of LMM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are



incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding LMM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding LMM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

Often the placement of a gene on the chromosome of another mammalian species, such as mouse,

may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, LMM, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between LMM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with LMM, or fragments thereof, and washed. Bound LMM is then detected by methods well known in the art. Purified LMM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding LMM specifically compete with a test compound for binding LMM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with LMM.

In additional embodiments, the nucleotide sequences which encode LMM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/216,801, U.S. Ser. 60/218,233, No. U.S. Ser. No. 60/220,046, U.S. Ser.

No. 60/220,739, U.S. Ser. No. 60/222,824, and U.S. Ser. No. 60/216,803 are expressly incorporated by reference herein.

## EXAMPLES

### 5 I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of  
10 phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated  
15 using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA  
20 libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the  
25 appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid  
30 (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

### II. Isolation of cDNA Clones

35 Plasmids obtained as described in Example I were recovered from host cells by in vivo

excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96  
5 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in  
10 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.  
15 Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as  
20 the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading  
25 frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and  
30 programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family  
databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary  
35 structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.

Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA  
5 assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length  
10 polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide  
15 and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and  
20 threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value,  
25 the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:9-16. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### 30 IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative lipid metabolism molecules were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and  
35 S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to

form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode lipid metabolism molecules, the encoded polypeptides were  
5 analyzed by querying against PFAM models for lipid metabolism molecules. Potential lipid metabolism molecules were also identified by homology to Incyte cDNA sequences that had been annotated as lipid metabolism molecules. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbprl public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to  
10 correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with  
15 Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

20 Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information,  
25 generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated  
30 but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent  
35 type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared

by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

5 Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in  
10 Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences  
15 were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of LMM Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:9-16 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other  
20 implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:9-16 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences  
25 had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between  
30 chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site  
35 (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified

disease genes map within or in proximity to the intervals indicated above.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding LMM are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous



system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding LMM. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### VIII. Extension of LMM Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by

electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and  
5 sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site  
10 overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following  
15 parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing  
20 primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

## 25 IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:9-16 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06  
30 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based  
35 hybridization analysis of human genomic DNA digested with one of the following endonucleases:

Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### **X. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40

$\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one  
5 with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is  
10 then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### **Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification  
15 uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope  
20 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

25 Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).  
30 Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### **Hybridization**

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### 10 **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a

linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

5 A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### **XI. Complementary Polynucleotides**

10 Sequences complementary to the LMM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring LMM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of LMM. To inhibit transcription, a  
15 complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the LMM-encoding transcript.

#### **XII. Expression of LMM**

Expression and purification of LMM is achieved using bacterial or virus-based expression  
20 systems. For expression of LMM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).  
25 Antibiotic resistant bacteria express LMM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of LMM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding LMM by either homologous recombination or bacterial-mediated  
30 transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther.  
35 7:1937-1945.)

In most expression systems, LMM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from LMM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified LMM obtained by these methods can be used directly in the assays shown in Examples XVI and XVII where applicable.

### XIII. Functional Assays

LMM function is assessed by expressing the sequences encoding LMM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of LMM on gene expression can be assessed using highly purified populations

of cells transfected with sequences encoding LMM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success  
5 NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding LMM and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### **XIV. Production of LMM Specific Antibodies**

LMM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g.,  
10 Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the LMM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for  
15 selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to  
20 increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-LMM activity by, for example, binding the peptide or LMM to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **25 XV. Purification of Naturally Occurring LMM Using Specific Antibodies**

Naturally occurring or recombinant LMM is substantially purified by immunoaffinity chromatography using antibodies specific for LMM. An immunoaffinity column is constructed by covalently coupling anti-LMM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is  
30 blocked and washed according to the manufacturer's instructions.

Media containing LMM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of LMM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/LMM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as  
35 urea or thiocyanate ion), and LMM is collected.



**XVI. Identification of Molecules Which Interact with LMM**

LMM, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled LMM, washed,  
5 and any wells with labeled LMM complex are assayed. Data obtained using different concentrations of LMM are used to calculate values for the number, affinity, and association of LMM with the candidate molecules.

Alternatively, molecules interacting with LMM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially  
10 available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

LMM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

**15 XVII. Demonstration of LME Activity**

LME activity can be demonstrated by an *in vitro* hydrolysis assay with vesicles containing 1-palmitoyl-2-[1- $^{14}\text{C}$ ]oleoyl phosphatidylcholine (Sigma-Aldrich). LME triglyceride lipase activity and phospholipase  $\text{A}_2$  activity are demonstrated by analysis of the cleavage products isolated from the hydrolysis reaction mixture.

20 Vesicles containing 1-palmitoyl-2-[1- $^{14}\text{C}$ ]oleoyl phosphatidylcholine (Amersham Pharmacia Biotech.) are prepared by mixing 2.0  $\mu\text{Ci}$  of the radiolabeled phospholipid with 12.5 mg of unlabeled 1-palmitoyl-2-oleoyl phosphatidylcholine and drying the mixture under  $\text{N}_2$ . 2.5 ml of 150 mM Tris-HCl, pH 7.5, is added, and the mixture is sonicated and centrifuged. The supernatant may be stored at 4  $^\circ\text{C}$ . The final reaction mixtures contain 0.25 ml of Hanks buffered salt solution  
25 supplemented with 2.0 mM taurochenodeoxycholate, 1.0% bovine serum albumin, 1.0 mM  $\text{CaCl}_2$ , pH 7.4, 150  $\mu\text{g}$  of 1-palmitoyl-2-[1- $^{14}\text{C}$ ]oleoyl phosphatidylcholine vesicles, and various amounts of LME diluted in PBS. After incubation for 30 min at 37  $^\circ\text{C}$ , 20  $\mu\text{g}$  each of lyso-phosphatidylcholine and oleic acid are added as carriers and each sample is extracted for total lipids. The lipids are  
30 separated by thin layer chromatography using a two solvent system of chloroform:methanol:acetic acid:water (65:35:8:4) until the solvent front is halfway up the plate. The process is then continued with hexane:ether:acetic acid (86:16:1) until the solvent front is at the top of the plate. The lipid-containing areas are visualized with  $\text{I}_2$  vapor; the spots are scraped, and their radioactivity is determined by scintillation counting. The amount of radioactivity released as fatty acids will increase as a greater amount of LME is added to the assay mixture while the amount of radioactivity released  
35 as lyso-phosphatidylcholine will remain low. This demonstrates that LME cleaves at the sn-2 and not

the sn-1 position, as is characteristic of phospholipase A<sub>2</sub> activity.

Alternatively, LME phospholipase activity is measured by the hydrolysis of a fatty acyl residue at the sn-1 position of phosphatidylserine. LME is combined with the Tritium [<sup>3</sup>H] labeled substrate phosphatidylserine at stoichiometric quantities in a suitable buffer. Following an appropriate incubation time, the hydrolyzed reaction products are separated from the substrates by chromatographic methods. The amount of acylglycerophosphoserine produced is measured by counting tritiated product with the help of a scintillation counter. Various control groups are set up to account for background noise and unincorporated substrate. The final counts represent the tritiated enzyme product [<sup>3</sup>H]-acylglycerophosphoserine, which is directly proportional to the activity of LME in biological samples.

LME lipoxigenase activity can be measured by chromatographic methods. Extracted LME lipoxigenase protein is incubated with 100 μM [1-<sup>14</sup>C] arachidonic acid or other unlabeled fatty acids at 37°C for 30 min. After the incubation, stop solution (acetonitrile:methanol:water, 350:150:1) is added. The samples are extracted and analyzed by reverse-phase HPLC by using a solvent system of methanol/water/acetic acid, 85:15:0.01 (vol/vol) at a flow rate of 1 ml/min. The effluent is monitored at 235 nm and analyzed for the presence of the major arachidonic metabolite such as 12-HPETE (catalyzed by 12-LOX). The fractions are also subjected to liquid scintillation counting. The final counts represent the products, which is directly proportional to the activity of LME in biological samples. For stereochemical analysis, the metabolites of arachidonic acid are analyzed further by chiral phase-HPLC and by mass spectrometry (Sun, D. et al. (1998) J. Biol. Chem. 273:33540-33547).

LMM activity of SEQ ID NO:7 can be determined, for example, by measuring the rate of incorporation of a radioactive fatty acid precursor into fatty acyl-CoA. The final reaction contains 200 mM Tris-HCl, pH 7.5, 2.5 mM ATP, 8 mM MgCl<sub>2</sub>, 2mM EDTA, 20 mM NaF, 0.1% Triton X-100, 10 mM [<sup>3</sup>H]oleate, [<sup>3</sup>H]myristate or [<sup>14</sup>C]decanoate, 0.5 mM coenzyme A, and LMM in a total volume of 0.5 ml. The reaction is initiated with the addition of coenzyme A, incubated at 35 °C for 10 min, and terminated by the addition of 2.5 ml of isopropyl alcohol, n-heptane, 1 M H<sub>2</sub> SO<sub>4</sub> (40:10:1). Radioactive fatty acid is removed by organic extraction using n-heptane. Fatty acyl-CoA formed during the reaction remains in the aqueous fraction and is quantified by scintillation counting (Black, P.N. et al. (1997) J. Biol. Chem. 272: 4896-4904).

In the alternative, degradation of the sphingolipid glucosylceramide can be used to measure LMM activity of SEQ ID NO:8. 25-50 microunits glucocerebrosidase are incubated with varying concentrations of LMM in a 40 μl reaction at 37 °C for 20 min. The final reaction contains 50mM sodium citrate pH 4.5, 20 ng human serum albumin, and 3.125 mM lipids in the form of liposomes, which contain lipids in the following proportions: [<sup>14</sup>C]glucosylceramide (3 mol %, 2.4 Ci/mol),

cholesterol (23 mol %), phosphatidic acid (20 mol%), phosphatidylcholine (54 mol %). The reaction is stopped by the addition of 160 µl chloroform/methanol (2:1) and 20 µl 0.1 % glucose, and shaking. After centrifugation at 4000 rpm, enzymatically released [<sup>14</sup>C]glucose in the aqueous phase is measured in a scintillation counter. LMM activity is determined by its effect on increasing the rate of

5 glucosylceramide hydrolysis by glucocerebrosidase (Wilkening, G. et al. J. Biol. Chem. (1998) 273:30271-30278).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the

10 invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

| Incyte<br>Project ID | Polypeptide<br>SEQ ID NO: | Incyte<br>Polypeptide ID | Polynucleotide<br>SEQ ID NO: | Incyte<br>Polynucleotide ID |
|----------------------|---------------------------|--------------------------|------------------------------|-----------------------------|
| 2181310              | 1                         | 2181310CD1               | 9                            | 2181310CB1                  |
| 2965233              | 2                         | 2965233CD1               | 10                           | 2965233CB1                  |
| 1281946              | 3                         | 1281946CD1               | 11                           | 1281946CB1                  |
| 2970737              | 4                         | 2970737CD1               | 12                           | 2970737CB1                  |
| 5924878              | 5                         | 5924878CD1               | 13                           | 5924878CB1                  |
| 7477093              | 6                         | 7477093CD1               | 14                           | 7477093CB1                  |
| 2194717              | 7                         | 2194717CD1               | 15                           | 2194717CB1                  |
| 7473574              | 8                         | 7473574CD1               | 16                           | 7473574CB1                  |

Table 2

| Polypeptide<br>SEQ ID NO: | Incyte<br>Polypeptide<br>ID | GenBank ID<br>NO: | Probability<br>score | GenBank Homolog  |
|---------------------------|-----------------------------|-------------------|----------------------|--|
| 1                         | 2181310CD1                  | g571466           | 3.6e-204             | [fl][Rattus norvegicus] phospholipase C delta-4<br>Lee, S.B. and Rhee, S.G. (1996) J. Biol. Chem. 271: 25-31   |
| 2                         | 2965233CD1                  | g6705987          | 1.7e-101             | [Mus musculus] phospholipase C-L2<br>Otsuki, M. et al. (1999) Biochem. Biophys. Res. Commun.<br>266: 97-103  |
| 3                         | 1281946CD1                  | g3811347          | 4.0e-156             | [fl][Homo sapiens] cytosolic phospholipase A2 beta<br>Pickard, R.T. et al. (1999) J. Biol. Chem. 274:8823-8831   |
| 4                         | 2970737CD1                  | g2565396          | 2.9e-74              | [fl][Mus musculus] schwannoma-associated protein   |
| 5                         | 5924878CD1                  | g2408127          | 5.8e-17              | [fl][Trypanosoma cruzi] glycosylphosphatidylinositol-<br>specific phospholipase C<br>Redpath, M. et al. (1998) Conservation of genetic<br>linkage between heat shock protein 100 and<br>glycosylphosphatidylinositol-specific phospholipase C<br>in Trypanosoma brucei and Trypanosoma cruzi Mol.<br>Biochem. Parasitol. 94:113-21 |
| 6                         | 7477093CD1                  | g3811347          | 1.5e-205             | [fl][Homo sapiens] cytosolic phospholipase A2 beta<br>Pickard, R.T. et al. (1999) J. Biol. Chem. 274: 8823-<br>8831  |
| 7                         | 2194717CD1                  | g1929067          | 8.90e-58             | fadD36 (a long chain fatty acid-CoA ligase homolog)<br>[Mycobacterium tuberculosis]  |
| 8                         | 7473574CD1                  | g337756           | 7.30e-89             | sphingolipid activator precursor [Homo sapiens]<br>Harzer, K. et al. (1989) Eur. J. Pediatr. 149:31-39   |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites   | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases  |
|------------|-----------------------|---------------------|---|-------------------------------|--|---|
| 1          | 2181310CD1            | 731                 | S100 S108 S155<br>S219 S286 S31<br>S391 S432 S454<br>S473 S491 S56<br>S618 S62 S64<br>S679 S710 S80<br>T11 T127 T194<br>T324 T379 T407<br>T436 T635 T68 |                               | Phosphatidylinositol-specific phospholipase<br>PI-PLC-X:E181-K192 D259-K405<br>PI-PLC-Y:A461-R578<br>C2 domain:<br>L598-T688<br>PH (pleckstrin homology) domain:<br>L17-D124<br>Phosphatidylinositol-specific phospholipase X-box domain proteins<br>BL50007B:T324-Q361<br>BL50007C:L389-K405<br>BL50007D:H511-G552<br>BL50007E:Y675-L711<br>Phospholipase C signature<br>PR00390A:P263-Q281<br>PR00390B:G290-G310<br>PR00390C:Q388-K405<br>PR00390D:L516-W537<br>PR00390E:W537-M555<br>PR00390F:L689-R699<br>C2 domain signature<br>PR00360A:E616-I628<br>PR00360B:N646-R659<br>PR00360C:L668-D676<br>Ef_Hand D147-V159<br>PHOSPHOLIPASE C PHOSPHODIESTERASE<br>HYDROLASE 1 PHOSPHATIDYLINOSITOL 4 5<br>BISPHOSPHATE LIPID DEGRADATION<br>PHOSPHOINOSITIDE SPECIFIC<br>PD001214:D259-T407<br>PD001202:L462-R578 | HMMER_PFAM<br>HMMER_PFAM<br>HMMER_PFAM<br>BLIMPS_BLOCKS<br>BLIMPS_PRINTS<br>BLIMPS_PRINTS<br>MOTIFS<br>BLAST_PRODUM |



Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites  | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs   | Analytical Methods and Databases  |
|------------|-----------------------|---------------------|--|-------------------------------|---|---|
| 2          | 2965233CD1            | 621                 | S125 S155 S187<br>S236 S277 S351<br>S446 S484 S488<br>S544 S551 S559<br>S566 S578 S585<br>S592 S619 S80<br>S92 T135 T174<br>T237 T388 T505<br>T513 | N291 N304<br>N473 N535        | Signal peptide:M1-R41<br>Phosphatidylinositol-specific phospholipase<br>PI-PLC-X:D324-K469<br>EF hand:W170-L198, R206-M235<br>PH domain:A45-A152<br>Phosphatidylinositol-specific phospholipase X-box domain<br>BL50007A:L329-G374<br>BL50007B:T388-Q425<br>BL50007C:L453-K469<br>Phospholipase C signature<br>PR00390A:P328-Q346<br>PR00390B:D354-G374<br>PR00390C:T452-K469<br>Ef_Hand:D179-V191<br>PHOSPHOLIPASE C PHOSPHODIESTERASE<br>HYDROLASE 1 PHOSPHATIDYLINOSITOL 4 5<br>BISPHOSPHATE LIPID DEGRADATION<br>PHOSPHOINOSITIDE SPECIFIC<br>PD001214:D324-K469<br>PHOSPHOLIPASE 1 PHOSPHATIDYLINOSITOL 4 5<br>BISPHOSPHATE PHOSPHODIESTERASE HYDROLASE<br>LIPID DEGRADATION C CALCIUMBINDING<br>PD004439:Q101-Q323, C42-Q265<br>1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE<br>PHOSPHODIESTERASE D<br>DM00855 P51178 64-472:S89-D495<br>DM00855 P08487 71-500:I88-G501<br>DM00855 P16885 63-486:I88-E486<br>DM00855 P40977 208-616:I90-N500 | SPSCAN<br>HMME PFAM<br>HMME PFAM<br>HMME PFAM<br>BLIMPS_BLOCKS<br>BLIMPS_PRINTS<br>MOTIFS<br>BLAST_PRODUM<br>BLAST_PRODUM<br>BLAST_DOMO |



Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites   | Potential Glycosylation Sites                                       | Signature Sequences, Domains and Motifs   | Analytical Methods and Databases  |
|------------|-----------------------|---------------------|---|---|---|---|
| 3          | 1281946CD1            | 848                 | S134 S242 S308<br>S321 S428 S457<br>S506 S566 S580<br>S63 S90 T113<br>T188 T268 T40<br>T436 T565 T626<br>T659 T718 T768<br>T781 T828 Y115 | N504 N590<br>N636 N766<br>N803 N88<br>N95 N504                      | Lysophospholipase catalytic domain (PLA2_B):<br>S279-A848<br>C2 domain:<br>L46-L129<br>CYTOSOLIC PHOSPHOLIPASE A2<br>PD014471:G396-S577, F635-N766, F571-T638<br>C2-DOMAIN<br>DM00150 P50393 4-128:V48-N148<br>DM00150 P47713 4-128:V48-N148<br>DM00150 B39898 4-128:V48-N148<br>Signal peptide: M1-S25<br>Signal peptide: M1-S25<br>Transmembrane domain: K5-F24<br>Phospholipase D active site motif PLDC: N153-S180, F372-D398<br>Phosphatidylinositol-specific phospholipase PI-PLC: T79-G167<br>Signal cleavage: M1-D38<br>C2 domain:<br>L232-L316<br>Lysophospholipase catalytic domain: P454-A1019<br>C2 domain signature<br>PR00360B:D274-L287<br>PR00360C:L296-D304<br>CYTOSOLIC PHOSPHOLIPASE A2 CPLA2: PD014471:G570-L741, E783-N942 | HMMER-PFAM<br>HMMER-PFAM<br>BLAST_PRODOR<br>BLAST_PRODOR<br>HMMER<br>SPSCAN<br>HMMER<br>HMMER-PFAM<br>HMMER-PFAM<br>SPSCAN<br>HMMER-PFAM<br>HMMER-PFAM<br>BLIMPS_PRINTS<br>BLAST_PRODOR |
| 4          | 2970737CD1            | 474                 | S210 S222 S25<br>S251 S269 S292<br>S313 S336 S348<br>S38 S458 S90<br>T293 T461 Y146<br>Y433   | N148 N193<br>N240 N259<br>N346 N369<br>N417 N447<br>N459 N59<br>N95 |   | HMMER<br>SPSCAN<br>HMMER<br>HMMER-PFAM  |
| 5          | 5924878CD1            | 321                 | S100 S118 S180<br>S232 S241 T18<br>T220 T234 T250<br>T267 T269 T282<br>Y111 Y189  | N218 N83  |   | HMMER-PFAM<br>HMMER-PFAM  |
| 6          | 7477093CD1            | 1026                |   |   |   | HMMER-PFAM  |

Table 3 (cont.)

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites                  | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs  | Analytical Methods and Databases   |
|------------|-----------------------|---------------------|--|-------------------------------|--|--|
| 7          | 2194717CD1            | 379                 | S288 S301 S332<br>T153 T275 T324<br>T36 T87 T9   | N148                          | AMP-binding enzyme M1-V307<br>Putative AMP-binding domain signature<br>BL00455: Y4-H19<br>AMP-BINDING SIGNATURE PR00154: M1-T8, T9-S17<br>LIGASE SYNTHETASE PROTEIN ENZYME<br>BIOSYNTHESIS ANTIBIOTIC<br>PHOSPHOPANTHETINE MULTIFUNCTIONAL<br>REPEAT ACYLCOA PD000070: I2-I308<br>PUTATIVE AMP-BINDING DOMAIN<br>DM00073 P38137 58-529: M1-N196, D211-L372<br>Putative AMP-binding domain signature<br>amp_binding.prf: I2-W35<br>Amp_Binding: I2-K13<br>signal cleavage: M1-G52<br>SAPOSIN REPEAT DM02041 P07602 84-254:D86-I195<br>signal peptide:M1-P19<br>signal cleavage:M1-A17 | HMMER_FFAM<br>BLIMPS_BLOCKS<br>BLIMPS_PRINTS<br>BLAST_PRODOM<br>BLAST_DOMO<br>PROFILESCAN<br>MOTIFS<br>SPScan<br>BLAST_DOMO<br>HMMER<br>SPScan |
| 8          | 7473574CD1            | 362                 | S138 S154 S255<br>S293 T149 T256<br>T284 T58 T83 |                               |  |  |

Table 4

| Polynucleotide<br>SEQ ID NO: | Incyte<br>Polynucleotide ID | Sequence<br>Length | Selected<br>Fragment(s)                         | Sequence Fragments                 | 5' Position | 3' Position |
|------------------------------|-----------------------------|--------------------|---|------------------------------------|-------------|-------------|
| 9                            | 2181310CB1                  | 3594               | 2635-2675,<br>817-1663,<br>1-78, 3079-<br>3594  | GRI. ENST00000045865.re<br>p.edit2 | 296         | 2469        |
|                              |                             |                    |   | 60126275D2                         | 1153        | 1358        |
|                              |                             |                    |   | 3349205H1 (BRAITUT24)              | 1           | 282         |
|                              |                             |                    |   | 6819982J1 (OVARDIR01)              | 2500        | 3191        |
|                              |                             |                    |   | 1598847F6 (BLADNOT03)              | 1518        | 2011        |
|                              |                             |                    |   | 727643H1 (SYNOAT01)                | 1338        | 1554        |
|                              |                             |                    |   | 1597679F6 (BRAINOT14)              | 1902        | 2471        |
|                              |                             |                    |   | 7231776H1 (BRAXTDR15)              | 2066        | 2701        |
|                              |                             |                    |   | 7680564H1 (BRAFTUE01)              | 2965        | 3594        |
|                              |                             |                    |   | 7930120H1 (COLNDIS02)              | 100         | 624         |
| 10                           | 2965233CB1                  | 2243               | 1970-2028,<br>1-761                             | 55047576H1                         | 1759        | 2243        |
|                              |                             |                    |   | 7097553F6 (BRACDIR02)              | 714         | 1606        |
|                              |                             |                    |   | 55058502J1                         | 435         | 1314        |
|                              |                             |                    |   | 6805811F6 (SKIRNOR01)              | 1           | 588         |
|                              |                             |                    |   | 55047574H1                         | 1540        | 2226        |
| 11                           | 1281946CB1                  | 2547               | 1-113, 639-<br>1814, 353-<br>479, 1973-<br>2467 | 7691008J1 (PROSTME06)              | 1211        | 1700        |
|                              |                             |                    |   | 70796816V1                         | 1770        | 2431        |
|                              |                             |                    |   | 70924316V1                         | 1966        | 2547        |
|                              |                             |                    |   | 7625880J1 (KIDNFEE02)              | 1587        | 2225        |
|                              |                             |                    |   | 70759625V1                         | 695         | 1350        |
| 12                           | 2970737CB1                  | 2673               | 1-454, 553-<br>1528                             | 5721842F6 (SEMVN0T05)              | 600         | 1284        |
|                              |                             |                    |   | 8102436H1 (EYERNOA01)              | 1           | 625         |
|                              |                             |                    |   | 70762139V1                         | 1383        | 2055        |
|                              |                             |                    |   | 70762131V1                         | 1249        | 1899        |

Table 4 (cont.)

| Polynucleotide<br>SEQ ID NO: | Incyte<br>Polynucleotide ID | Sequence<br>Length | Selected<br>Fragment(s)                       | Sequence Fragments               | 5' Position | 3' Position |
|------------------------------|-----------------------------|--------------------|---|----------------------------------|-------------|-------------|
| 12                           |                             |                    |   | 71082605V1                       | 2081        | 2673        |
|                              |                             |                    |   | 2708432T6 (PONSAT01)             | 1959        | 2653        |
| 13                           | 5924878CB1                  | 1422               | 1393-1422,<br>1-67, 948-<br>1015, 588-<br>607 | 55072677J1                       | 1           | 146         |
|                              |                             |                    |   | 5956411F6 (BRATNOT05)            | 13          | 735         |
|                              |                             |                    |   | GNN.G6939631_000017_00<br>2      | 332         | 1041        |
|                              |                             |                    |   | 5924878T8 (BRAIFET02)            | 751         | 1422        |
| 14                           | 7477093CB1                  | 3197               | 1-1504,<br>2045-2067,<br>2212-3197            | 2025787H1 (KERANOT02)            | 2970        | 3197        |
|                              |                             |                    |   | GNN.G6587989_004.edit            | 1           | 1159        |
|                              |                             |                    |   | 3095305F6 (CERVNOT03)            | 1494        | 1593        |
|                              |                             |                    |   | GNN.G8705173_000028_00<br>2.edit | 671         | 3081        |
|                              |                             |                    |   | 6501441H1 (PROSTUS25)            | 2169        | 2861        |
|                              |                             |                    |   | 3363958H1 (PROSBPT02)            | 538         | 723         |
| 15                           | 2194717CB1                  | 2446               | 2427-2446                                     | 1860248F6 (PROSNOT18)            | 1496        | 2089        |
|                              |                             |                    |   | 70092485V1                       | 1994        | 2446        |
|                              |                             |                    |   | 6999602H1 (HEALDIR01)            | 898         | 1582        |
|                              |                             |                    |   | 70784306V1                       | 806         | 1444        |
|                              |                             |                    |   | 2194717F6 (THYRTUT03)            | 1576        | 2093        |
|                              |                             |                    |   | 4704969F6 (SMCRXT01)             | 202         | 648         |
|                              |                             |                    |   | 70782463V1                       | 351         | 848         |
|                              |                             |                    |   | 60217609D1                       | 1           | 339         |
| 16                           | 7473574CB1                  | 2097               | 1-273, 334-<br>693, 1011-<br>1516             | 7258126H1 (SKIRTD01)             | 1272        | 1670        |
|                              |                             |                    |   | 5612966F8 (SKINTDT01)            | 1007        | 1616        |
|                              |                             |                    |   | FL890761_00001                   | 1           | 1260        |
|                              |                             |                    |   | g2110818                         | 1517        | 2097        |

Table 5

| Polynucleotide<br>SEQ ID NO: | Project ID | Representative Library |
|------------------------------|------------|------------------------|
| 9                            | 2181310CB1 | PGANNT01               |
| 10                           | 2965233CB1 | SKIRNOR01              |
| 11                           | 1281946CB1 | SKIRNOR01              |
| 12                           | 2970737CB1 | BRADDIR01              |
| 13                           | 5924878CB1 | BRAIFET02              |
| 14                           | 7477093CB1 | PROSTUS25              |
| 15                           | 2194717CB1 | FIBRTXS07              |
| 16                           | 7473574CB1 | SKINTDT01              |

Table 6

| Library   | Vector  | Library Description  |
|-----------|---------|--|
| BRADDIR01 | pINCY   | Library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.   |
| BRAIFET02 | pINCY   | Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation.  |
| FIBRTXS07 | pINCY   | This subtracted library was constructed using 1.3 million clones from a dermal fibroblast library and was subjected to two rounds of subtraction hybridization with 2.8 million clones from the an untreated dermal fibroblast tissue library. The starting library for subtraction was constructed using RNA isolated from treated dermal fibroblast tissue removed from the breast of a 31-year-old Caucasian female. The cells were treated with 9CIS retinoic acid. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated dermal fibroblast tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.  |
| PGANNOT01 | PSPORT1 | Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule. Surgical margins were negative for tumor.  |
| PROSTUS25 | pINCY   | This subtracted prostate tumor tissue library was constructed using 2.36 million clones from a prostate tumor tissue library and was subjected to two rounds of subtraction hybridization with 2.36 million clones from an untreated prostate epithelial cell tissue library. The starting library for subtraction was constructed using RNA isolated from prostate tumor tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3) involving the left and right sides of the prostate peripherally with invasion of the capsule. Adenofibromatous hyperplasia was present. Patient history included elevated prostate-specific antigen (PSA), diverticulitis of colon, asbestosis, and thrombophlebitis. Family history included benign hypertension multiple myeloma, hyperlipidemia, and rheumatoid arthritis. The hybridization probe for subtraction was derived from a similarly constructed library using RNA isolated from untreated prostate epithelial cell tissue from a different donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791. |

Table 6 (cont.)

| Library   | Vector   | Library Description  |
|-----------|----------|--|
| SKINTDT01 | PINCY    | Library was constructed using two pooled libraries. The first library was constructed using RNA isolated from breast skin tissue removed from a 46-year-old Caucasian female during breast biopsy and unilateral extended simple mastectomy. Pathology for the non-tumorous breast tissue indicated mildly proliferative fibrocystic changes. Pathology for the associated breast tumor tissue indicated intraductal carcinoma and multifocal ductal carcinoma in situ, both comedo and non comedo types with extensive intraductal calcifications. Patient history included deficiency anemia, chronic sinusitis, extrinsic asthma, kidney infection, and a normal delivery. Family history included diabetes type II, benign hypertension, cerebrovascular accident, malignant neoplasm of the skin, and hyperlipidemia. The second library was constructed in the same manner using RNA isolated from the skin of a 20-week-old Caucasian fetus who died from Patau's Syndrome. |
| SKIRNOR01 | PCDNA2.1 | This random primed library was constructed using RNA isolated from skin tissue removed from the breast of a 17-year-old Caucasian female during bilateral reduction mammoplasty. Patient history included breast hypertrophy. Family history included benign hypertension.   |

Table 7

| Program           | Description   | Reference  | Parameter Threshold   |
|-------------------|---|--|---|
| ABI FACTURA       | A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.  | Applied Biosystems, Foster City, CA.   |   |
| ABI/PARACEL FDF   | A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.   | Applied Biosystems, Foster City, CA;<br>Paracel Inc., Pasadena, CA.  | Mismatch <50%   |
| ABI AutoAssembler | A program that assembles nucleic acid sequences.  | Applied Biosystems, Foster City, CA.   |   |
| BLAST             | A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.                    | Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.  | ESTs: Probability value= 1.0E-8 or less<br>Full Length sequences: Probability value= 1.0E-10 or less  |
| FASTA             | A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch. | Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.              | ESTs: fasta E value=1.06E-6<br>Assembled ESTs: fasta Identity= 95% or greater and<br>Match length=200 bases or greater;<br>fastx E value=1.0E-8 or less<br>Full Length sequences:<br>fastx score=100 or greater |
| BLIMPS            | A BLOCKS IMPROVED Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.            | Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424. | Probability value= 1.0E-3 or less   |
| HMMER             | An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.  | Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.   | PFAM hits: Probability value= 1.0E-3 or less<br>Signal peptide hits: Score= 0 or greater  |



Table 7 (cont.)

| Program     | Description   | Reference  | Parameter Threshold  |
|-------------|---|--|--|
| ProfileScan | An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.   | Gribskov, M. et al. (1988) CABIOS 4:61-66;<br>Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.                                    | Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif.<br>Generally, score=1.4-2.1. |
| Phred       | A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.  | Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.  |  |
| Phrap       | A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences. | Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.             | Score= 120 or greater;<br>Match length= 56 or greater  |
| Consed      | A graphical tool for viewing and editing Phrap assemblies.  | Gordon, D. et al. (1998) Genome Res. 8:195-202.  |  |
| SPScan      | A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.  | Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.  | Score=3.5 or greater   |
| TMAP        | A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.   | Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.   |  |
| TMHMMER     | A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.   | Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182. |  |
| Motifs      | A program that searches amino acid sequences for patterns that matched those defined in Prosite.  | Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.   |  |

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of  
5 SEQ ID NO:1-8,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical  
to an amino acid sequence selected from the group consisting of SEQ ID NO:1-3 and SEQ ID NO:5-  
8,
  - c) a polypeptide comprising a naturally occurring amino acid sequence at least 80% identical  
10 to an amino acid sequence of SEQ ID NO:4,
  - d) a biologically active fragment of a polypeptide having an amino acid sequence selected  
from the group consisting of SEQ ID NO:1-8, and
  - e) an immunogenic fragment of a polypeptide having an amino acid sequence selected from  
the group consisting of SEQ ID NO:1-8.
- 15 2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-8.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
- 20 4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID  
NO:9-16.
- 25 6. A recombinant polynucleotide comprising a promoter sequence operably linked to a  
polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
- 30 8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method for producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said  
cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide  
35 comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of

claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

5

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-11 and SEQ ID NO:13-16,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 80% identical to a polynucleotide sequence of SEQ ID NO:12,

d) a polynucleotide complementary to a polynucleotide of a),

15 e) a polynucleotide complementary to a polynucleotide of b),

f) a polynucleotide complementary to a polynucleotide of c), and

g) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

20

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

30

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction

35

amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

5           16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

10           17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

18. A method for treating a disease or condition associated with decreased expression of functional LMM, comprising administering to a patient in need of such treatment the composition of claim 16.

15           19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20           20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

25           21. A method for treating a disease or condition associated with decreased expression of functional LMM, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- 30           a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

35           24. A method for treating a disease or condition associated with overexpression of functional

LMM, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

- 5       a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

10       26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound,
- 15       and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

20

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions
- 25       suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

30       28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological
- 35       sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim

11 or fragment thereof;

- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of LMM in a biological sample comprising the steps of:

- a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of LMM in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of LMM in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim

10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

5 b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

10 36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

15 38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

20 c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

25

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

30 41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

5 b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

15 45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

20

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

25

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

30

53. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:9.

54. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:10.



55. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:11.

56. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
5 NO:12.

57. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:13.

10 58. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:14.

59. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:15.  
15

60. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID  
NO:16.

<110> INCYTE GENOMICS, INC.

TANG, Y. Tom  
 AZIMZAI, Yalda  
 DAS, Debopriya  
 THORNTON, Michael  
 LU, Dyung Aina M.  
 TRIBOULEY, Catherine M.  
 YUE, Henry  
 GANDHI, Ameena R.  
 WALIA, Narinder K.  
 KHAN, Farrah A.  
 LU, Yan  
 YAO, Monique G.  
 HAFALIA, April J. A.  
 ELLIOTT, Vicki S.  
 PATTERSON, Chandra  
 LAL, Preeti  
 RAMKUMAR, Jayalaxmi  
 NGUYEN, Danniel B.  
 BAUGHN, Mariah R.

<120> LIPID METABOLISM MOLECULES

<130> PI-0152 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/216,803; 60/216,801; 60/218,233; 60/220,046; 60/220,739; 60/222,824

<151> 2000-07-07; 2000-07-07; 2000-07-14; 2000-07-21; 2000-07-26; 2000-08-04

<160> 16

<170> PERL Program

<210> 1

<211> 731

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2181310CD1

<400> 1

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Ser | Leu | Leu | Gln | Asp | Gln | Leu | Thr | Thr | Asp | Gln | Asp | Leu |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |
| Leu | Leu | Met | Gln | Glu | Gly | Met | Pro | Met | Arg | Lys | Val | Arg | Ser | Lys |
|     |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |
| Ser | Trp | Lys | Lys | Leu | Arg | Tyr | Phe | Arg | Leu | Gln | Asn | Asp | Gly | Met |
|     |     |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |
| Thr | Val | Trp | His | Ala | Arg | Gln | Ala | Arg | Gly | Ser | Ala | Lys | Pro | Ser |
|     |     |     |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |
| Phe | Ser | Ile | Ser | Asp | Val | Glu | Thr | Ile | Arg | Asn | Gly | His | Asp | Ser |
|     |     |     |     | 65  |     |     |     |     | 70  |     |     |     |     | 75  |
| Glu | Leu | Leu | Arg | Ser | Leu | Ala | Glu | Glu | Leu | Pro | Leu | Glu | Gln | Gly |
|     |     |     |     | 80  |     |     |     |     | 85  |     |     |     |     | 90  |
| Phe | Thr | Ile | Val | Phe | His | Gly | Arg | Arg | Ser | Asn | Leu | Asp | Leu | Met |
|     |     |     |     | 95  |     |     |     |     | 100 |     |     |     |     | 105 |
| Ala | Asn | Ser | Val | Glu | Ala | Gln | Ile | Trp | Met | Arg | Gly | Leu | Gln |     |
|     |     |     |     | 110 |     |     |     |     | 115 |     |     |     |     | 120 |
| Leu | Leu | Val | Asp | Leu | Val | Thr | Ser | Met | Asp | His | Gln | Glu | Arg | Leu |
|     |     |     |     | 125 |     |     |     |     | 130 |     |     |     |     | 135 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asp | Gln | Trp | Leu | Ser | Asp | Trp | Phe | Gln | Arg | Gly | Asp | Lys | Asn | Gln |
|     |     |     |     | 140 |     |     |     |     |     | 145 |     |     |     | 150 |
| Asp | Gly | Lys | Met | Ser | Phe | Gln | Glu | Val | Gln | Arg | Leu | Leu | His | Leu |
|     |     |     |     | 155 |     |     |     |     |     | 160 |     |     |     | 165 |
| Met | Asn | Val | Glu | Met | Asp | Gln | Glu | Tyr | Ala | Phe | Ser | Leu | Phe | Gln |
|     |     |     |     | 170 |     |     |     |     |     | 175 |     |     |     | 180 |
| Glu | Leu | Arg | Arg | Lys | Ile | Leu | Val | Lys | Gly | Lys | Lys | Leu | Thr | Leu |
|     |     |     |     | 185 |     |     |     |     |     | 190 |     |     |     | 195 |
| Glu | Glu | Asp | Leu | Glu | Tyr | Glu | Glu | Glu | Glu | Ala | Glu | Pro | Glu | Leu |
|     |     |     |     | 200 |     |     |     |     |     | 205 |     |     |     | 210 |
| Glu | Glu | Ser | Glu | Leu | Ala | Leu | Glu | Ser | Gln | Phe | Glu | Thr | Glu | Pro |
|     |     |     |     | 215 |     |     |     |     |     | 220 |     |     |     | 225 |
| Gly | Lys | Leu | Arg | His | Val | Leu | Ser | Met | Asp | Gly | Phe | Leu | Ser | Tyr |
|     |     |     |     | 230 |     |     |     |     |     | 235 |     |     |     | 240 |
| Leu | Cys | Ser | Lys | Asp | Gly | Asp | Ile | Phe | Asn | Pro | Ala | Cys | Leu | Pro |
|     |     |     |     | 245 |     |     |     |     |     | 250 |     |     |     | 255 |
| Ile | Tyr | Gln | Asp | Met | Thr | Gln | Pro | Leu | Asn | His | Tyr | Phe | Ile | Cys |
|     |     |     |     | 260 |     |     |     |     |     | 265 |     |     |     | 270 |
| Ser | Ser | His | Asn | Thr | Tyr | Leu | Val | Gly | Asp | Gln | Leu | Cys | Gly | Gln |
|     |     |     |     | 275 |     |     |     |     |     | 280 |     |     |     | 285 |
| Ser | Ser | Val | Glu | Gly | Tyr | Ile | Arg | Arg | Ala | Leu | Lys | Arg | Gly | Cys |
|     |     |     |     | 290 |     |     |     |     |     | 295 |     |     |     | 300 |
| Arg | Cys | Val | Glu | Val | Asp | Val | Trp | Asp | Gly | Pro | Ser | Gly | Glu | Pro |
|     |     |     |     | 305 |     |     |     |     |     | 310 |     |     |     | 315 |
| Val | Val | Tyr | His | Gly | His | Thr | Leu | Thr | Ser | Arg | Ile | Leu | Phe | Lys |
|     |     |     |     | 320 |     |     |     |     |     | 325 |     |     |     | 330 |
| Asp | Val | Val | Ala | Thr | Val | Ala | Gln | Tyr | Ala | Phe | Gln | Thr | Ser | Asp |
|     |     |     |     | 335 |     |     |     |     |     | 340 |     |     |     | 345 |
| Tyr | Pro | Val | Ile | Leu | Ser | Leu | Glu | Thr | His | Cys | Ser | Trp | Glu | Gln |
|     |     |     |     | 350 |     |     |     |     |     | 355 |     |     |     | 360 |
| Gln | Gln | Thr | Met | Ala | Arg | His | Leu | Thr | Glu | Ile | Leu | Gly | Glu | Gln |
|     |     |     |     | 365 |     |     |     |     |     | 370 |     |     |     | 375 |
| Leu | Leu | Ser | Thr | Thr | Leu | Asp | Gly | Val | Leu | Pro | Thr | Gln | Leu | Pro |
|     |     |     |     | 380 |     |     |     |     |     | 385 |     |     |     | 390 |
| Ser | Pro | Glu | Glu | Leu | Arg | Arg | Lys | Ile | Leu | Val | Lys | Gly | Lys | Lys |
|     |     |     |     | 395 |     |     |     |     |     | 400 |     |     |     | 405 |
| Leu | Thr | Leu | Glu | Glu | Asp | Leu | Glu | Tyr | Glu | Glu | Glu | Glu | Ala | Glu |
|     |     |     |     | 410 |     |     |     |     |     | 415 |     |     |     | 420 |
| Pro | Glu | Leu | Glu | Glu | Ser | Glu | Leu | Ala | Leu | Glu | Ser | Gln | Phe | Glu |
|     |     |     |     | 425 |     |     |     |     |     | 430 |     |     |     | 435 |
| Thr | Glu | Pro | Glu | Pro | Gln | Glu | Gln | Asn | Leu | Gln | Asn | Lys | Asp | Lys |
|     |     |     |     | 440 |     |     |     |     |     | 445 |     |     |     | 450 |
| Lys | Lys | Lys | Ser | Lys | Pro | Ile | Leu | Cys | Pro | Ala | Leu | Ser | Ser | Leu |
|     |     |     |     | 455 |     |     |     |     |     | 460 |     |     |     | 465 |
| Val | Ile | Tyr | Leu | Lys | Ser | Val | Ser | Phe | Arg | Ser | Phe | Thr | His | Ser |
|     |     |     |     | 470 |     |     |     |     |     | 475 |     |     |     | 480 |
| Lys | Glu | His | Tyr | His | Phe | Tyr | Glu | Ile | Ser | Ser | Phe | Ser | Glu | Thr |
|     |     |     |     | 485 |     |     |     |     |     | 490 |     |     |     | 495 |
| Lys | Ala | Lys | Arg | Leu | Ile | Lys | Glu | Ala | Gly | Asn | Glu | Phe | Val | Gln |
|     |     |     |     | 500 |     |     |     |     |     | 505 |     |     |     | 510 |
| His | Asn | Thr | Trp | Gln | Leu | Ser | Arg | Val | Tyr | Pro | Ser | Gly | Leu | Arg |
|     |     |     |     | 515 |     |     |     |     |     | 520 |     |     |     | 525 |
| Thr | Asp | Ser | Ser | Asn | Tyr | Asn | Pro | Gln | Glu | Leu | Trp | Asn | Ala | Gly |
|     |     |     |     | 530 |     |     |     |     |     | 535 |     |     |     | 540 |
| Cys | Gln | Met | Val | Ala | Met | Asn | Met | Gln | Thr | Ala | Gly | Leu | Glu | Met |
|     |     |     |     | 545 |     |     |     |     |     | 550 |     |     |     | 555 |
| Asp | Ile | Cys | Asp | Gly | His | Phe | Arg | Gln | Asn | Gly | Gly | Cys | Gly | Tyr |
|     |     |     |     | 560 |     |     |     |     |     | 565 |     |     |     | 570 |
| Val | Leu | Lys | Pro | Asp | Phe | Leu | Arg | Asp | Ile | Gln | Ser | Ser | Phe | His |
|     |     |     |     | 575 |     |     |     |     |     | 580 |     |     |     | 585 |
| Pro | Glu | Lys | Pro | Ile | Ser | Pro | Phe | Lys | Ala | Gln | Thr | Leu | Leu | Ile |
|     |     |     |     | 590 |     |     |     |     |     | 595 |     |     |     | 600 |
| Gln | Val | Ile | Ser | Gly | Gln | Gln | Leu | Pro | Lys | Val | Asp | Lys | Thr | Lys |

|                 |                     |                     |     |  |     |
|-----------------|---------------------|---------------------|-----|--|-----|
|                 | 605                 |                     | 610 |  | 615 |
| Glu Gly Ser Ile | Val Asp Pro Leu Val | Lys Val Gln Ile Phe | Gly |  |     |
|                 | 620                 |                     | 625 |  | 630 |
| Val Arg Leu Asp | Thr Ala Arg Gln Glu | Thr Asn Tyr Val Glu | Asn |  |     |
|                 | 635                 |                     | 640 |  | 645 |
| Asn Gly Phe Asn | Pro Tyr Trp Gly Gln | Thr Leu Cys Phe Arg | Val |  |     |
|                 | 650                 |                     | 655 |  | 660 |
| Leu Val Pro Glu | Leu Ala Met Leu Arg | Phe Val Val Met Asp | Tyr |  |     |
|                 | 665                 |                     | 670 |  | 675 |
| Asp Trp Lys Ser | Arg Asn Asp Phe Ile | Gly Gln Tyr Thr Leu | Pro |  |     |
|                 | 680                 |                     | 685 |  | 690 |
| Trp Thr Cys Met | Gln Gly Tyr Arg     | His Ile His Leu Leu | Ser |  |     |
|                 | 695                 |                     | 700 |  | 705 |
| Lys Asp Gly Ile | Ser Leu Arg Pro Ala | Ser Ile Phe Val Tyr | Ile |  |     |
|                 | 710                 |                     | 715 |  | 720 |
| Cys Ile Gln Glu | Gly Leu Glu Gly Asp | Glu Ser             |     |  |     |
|                 | 725                 |                     | 730 |  |     |

&lt;210&gt; 2

&lt;211&gt; 621

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2965233CD1

&lt;400&gt; 2

|                 |                 |                 |             |
|-----------------|-----------------|-----------------|-------------|
| Met Met Ala Pro | Pro Thr Ala Gly | Pro Leu Pro Gly | Pro Ala Leu |
| 1               | 5               | 10              | 15          |
| Pro Pro Glu Asp | Pro Gly Pro Asp | Pro Glu Ser Arg | Trp Leu Phe |
|                 | 20              | 25              | 30          |
| Leu Ser Ala Asn | Ile Leu Pro Val | Val Glu Arg Cys | Met Gly Ala |
|                 | 35              | 40              | 45          |
| Met Gln Glu Gly | Met Gln Met Val | Lys Leu Arg Gly | Gly Ser Lys |
|                 | 50              | 55              | 60          |
| Gly Leu Val Arg | Phe Tyr Tyr Leu | Asp Glu His Arg | Ser Cys Ile |
|                 | 65              | 70              | 75          |
| Arg Trp Arg Pro | Ser Arg Lys Asn | Glu Lys Ala Lys | Ile Ser Ile |
|                 | 80              | 85              | 90          |
| Asp Ser Ile Gln | Glu Val Ser Glu | Gly Arg Gln Ser | Glu Val Phe |
|                 | 95              | 100             | 105         |
| Gln Arg Tyr Pro | Asp Gly Ser Phe | Asp Pro Asn Cys | Cys Phe Ser |
|                 | 110             | 115             | 120         |
| Ile Tyr His Gly | Ser His Arg Glu | Ser Leu Asp Leu | Val Ser Thr |
|                 | 125             | 130             | 135         |
| Ser Ser Glu Val | Ala Arg Thr Trp | Val Thr Gly Leu | Arg Tyr Leu |
|                 | 140             | 145             | 150         |
| Met Ala Gly Ile | Ser Asp Glu Asp | Ser Leu Ala Arg | Arg Gln Arg |
|                 | 155             | 160             | 165         |
| Thr Arg Asp Gln | Trp Leu Lys Gln | Thr Phe Asp Glu | Ala Asp Lys |
|                 | 170             | 175             | 180         |
| Asn Gly Asp Gly | Ser Leu Ser Ile | Gly Glu Val Leu | Gln Leu Leu |
|                 | 185             | 190             | 195         |
| His Lys Leu Asn | Val Asn Leu Pro | Arg Gln Arg Val | Lys Gln Met |
|                 | 200             | 205             | 210         |
| Phe Arg Glu Ala | Asp Thr Asp Asp | His Gln Gly Thr | Leu Gly Phe |
|                 | 215             | 220             | 225         |
| Glu Glu Phe Cys | Ala Phe Tyr Lys | Met Met Ser Thr | Arg Arg Asp |
|                 | 230             | 235             | 240         |
| Leu Tyr Leu Leu | Met Leu Thr Tyr | Ser Asn His Lys | Asp His Leu |
|                 | 245             | 250             | 255         |
| Asp Ala Ala Ser | Leu Gln Arg Phe | Leu Gln Val Glu | Gln Lys Met |

|                         |     |                         |     |
|-------------------------|-----|-------------------------|-----|
| Ala Gly Val Thr         | 260 | 265                     | 270 |
| Leu Glu Ser Cys Gln Asp |     | Ile Ile Glu Gln Phe     |     |
|                         | 275 | 280                     | 285 |
| Glu Pro Cys Pro         |     | Gly Leu Leu Gly Ile Asp |     |
|                         | 290 | 295                     | 300 |
| Gly Phe Thr Asn Tyr     |     | Ala Gly Asp Ile Phe Asn |     |
|                         | 305 | 310                     | 315 |
| Pro Glu His His         |     | Met Thr Gln Pro Leu Ser |     |
|                         | 320 | 325                     | 330 |
| His Tyr Phe Ile         |     | Thr Tyr Leu Val Gly Asp |     |
|                         | 335 | 340                     | 345 |
| Gln Leu Met Ser         |     | Met Tyr Ala Trp Val Leu |     |
|                         | 350 | 355                     | 360 |
| Gln Ala Gly Cys         |     | Asp Cys Trp Asp Gly Pro |     |
|                         | 365 | 370                     | 375 |
| Asp Gly Glu Pro         |     | Tyr Thr Leu Thr Ser Lys |     |
|                         | 380 | 385                     | 390 |
| Ile Leu Phe Lys         |     | Ile Asn Lys Tyr Ala Phe |     |
|                         | 395 | 400                     | 405 |
| Ile Lys Asn Glu         |     | Ser Ile Glu Asn His Cys |     |
|                         | 410 | 415                     | 420 |
| Ser Val Ile Gln         |     | Gln Tyr Leu Thr Asp Ile |     |
|                         | 425 | 430                     | 435 |
| Leu Gly Asp Lys         |     | Val Ser Ser Glu Asp Ala |     |
|                         | 440 | 445                     | 450 |
| Thr Thr Leu Pro         |     | Lys Gly Lys Ile Leu Val |     |
|                         | 455 | 460                     | 465 |
| Lys Gly Lys Lys         |     | Ser Glu Asp Ala Glu Glu |     |
|                         | 470 | 475                     | 480 |
| Gly Glu Val Ser         |     | Asp Glu Ile Asp Asp Asp |     |
|                         | 485 | 490                     | 495 |
| Cys Lys Leu Leu         |     | Thr Asn Arg Lys Arg Val |     |
|                         | 500 | 505                     | 510 |
| Glu Asn Thr Ala         |     | Ser Leu Ile Lys Glu Ser |     |
|                         | 515 | 520                     | 525 |
| Lys Ile Arg Asp         |     | Asn Phe Ser Val Ser Thr |     |
|                         | 530 | 535                     | 540 |
| Leu Ser Pro Ser         |     | Lys Ser Lys Ala Glu Glu |     |
|                         | 545 | 550                     | 555 |
| Asp Val Glu Ser         |     | Ala Ser Arg Arg Asn Gly |     |
|                         | 560 | 565                     | 570 |
| Arg Leu Val Val         |     | Arg Lys Lys Lys Gly Ser |     |
|                         | 575 | 580                     | 585 |
| Lys Leu Lys Lys         |     | Glu Gly Asp Glu Gly Gln |     |
|                         | 590 | 595                     | 600 |
| Asp Ser Pro Gly         |     | Ala Thr Arg Gln Lys Ser |     |
|                         | 605 | 610                     | 615 |
| Ser Val Pro Ser         |     | Thr Arg                 |     |
|                         | 620 |                         |     |

&lt;210&gt; 3

&lt;211&gt; 848

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1281946CD1

&lt;400&gt; 3

Met Leu Trp Ala Leu Trp Pro Arg Trp Leu Ala Asp Lys Met Leu

1

5

10

15

Pro Leu Leu Gly Ala Val Leu Leu Gln Lys Arg Glu Lys Arg Val

|                 |   |  |     |  |     |
|-----------------|---|--|-----|--|-----|
|                 | 20  |  | 25  |  | 30  |
| Pro Leu Trp Arg | His Trp Arg Arg Glu Thr Tyr Pro Tyr Tyr Asp |  |     |  |     |
|                 | 35  |  | 40  |  | 45  |
| Leu Gln Val Lys | Val Leu Arg Ala Thr Asn Ile Arg Gly Thr Asp |  |     |  |     |
|                 | 50  |  | 55  |  | 60  |
| Leu Leu Ser Lys | Ala Asp Cys Tyr Val Gln Leu Trp Leu Pro Thr |  |     |  |     |
|                 | 65  |  | 70  |  | 75  |
| Ala Ser Pro Ser | Pro Ala Gln Thr Arg Ile Val Ala Asn Cys Ser |  |     |  |     |
|                 | 80  |  | 85  |  | 90  |
| Asp Pro Glu Trp | Asn Glu Thr Phe His Tyr Gln Ile His Gly Ala |  |     |  |     |
|                 | 95  |  | 100 |  | 105 |
| Val Lys Asn Val | Leu Glu Leu Thr Leu Tyr Asp Lys Asp Ile Leu |  |     |  |     |
|                 | 110   |  | 115 |  | 120 |
| Gly Ser Asp Gln | Leu Ser Leu Leu Leu Phe Asp Leu Arg Ser Leu |  |     |  |     |
|                 | 125   |  | 130 |  | 135 |
| Lys Cys Gly Gln | Pro His Lys His Thr Phe Pro Leu Asn His Gln |  |     |  |     |
|                 | 140   |  | 145 |  | 150 |
| Asp Ser Gln Glu | Leu Gln Val Glu Phe Val Leu Glu Lys Ser Gln |  |     |  |     |
|                 | 155   |  | 160 |  | 165 |
| Val Pro Ala Ser | Glu Val Ile Thr Asn Gly Val Leu Val Ala His |  |     |  |     |
|                 | 170   |  | 175 |  | 180 |
| Pro Cys Leu Arg | Ile Gln Gly Thr Leu Arg Gly Asp Gly Thr Ala |  |     |  |     |
|                 | 185   |  | 190 |  | 195 |
| Pro Arg Glu Glu | Tyr Gly Ser Arg Gln Leu Gln Leu Ala Val Pro |  |     |  |     |
|                 | 200   |  | 205 |  | 210 |
| Gly Ala Tyr Glu | Lys Pro Gln Leu Leu Pro Leu Gln Pro Pro Thr |  |     |  |     |
|                 | 215   |  | 220 |  | 225 |
| Glu Pro Gly Leu | Pro Pro Thr Phe Thr Phe His Val Asn Pro Val |  |     |  |     |
|                 | 230   |  | 235 |  | 240 |
| Leu Ser Ser Arg | Leu His Val Glu Leu Met Glu Leu Leu Ala Ala |  |     |  |     |
|                 | 245   |  | 250 |  | 255 |
| Val Gln Ser Gly | Pro Ser Ala Glu Leu Glu Ala Gln Thr Ser Lys |  |     |  |     |
|                 | 260   |  | 265 |  | 270 |
| Leu Gly Glu Gly | Gly Ile Leu Leu Ser Ser Leu Pro Leu Gly Gln |  |     |  |     |
|                 | 275   |  | 280 |  | 285 |
| Glu Glu Gln Cys | Ser Val Ala Leu Gly Glu Gly Glu Gly Val Ala |  |     |  |     |
|                 | 290   |  | 295 |  | 300 |
| Leu Ser Met Thr | Val Glu Met Ser Ser Gly Asp Leu Asp Leu Arg |  |     |  |     |
|                 | 305   |  | 310 |  | 315 |
| Leu Gly Phe Asp | Leu Ser Asp Gly Glu Gln Glu Phe Leu Asp Arg |  |     |  |     |
|                 | 320   |  | 325 |  | 330 |
| Arg Lys Gln Val | Val Ser Lys Ala Leu Gln Gln Val Leu Gly Leu |  |     |  |     |
|                 | 335   |  | 340 |  | 345 |
| Ser Glu Ala Leu | Asp Ser Gly Gln Val Pro Val Val Ala Val Leu |  |     |  |     |
|                 | 350   |  | 355 |  | 360 |
| Gly Ser Gly Gly | Gly Thr Arg Ala Met Ser Ser Leu Tyr Gly Ser |  |     |  |     |
|                 | 365   |  | 370 |  | 375 |
| Leu Ala Gly Leu | Gln Glu Leu Gly Leu Leu Asp Thr Val Thr Tyr |  |     |  |     |
|                 | 380   |  | 385 |  | 390 |
| Leu Ser Gly Val | Ser Gly Ser Thr Trp Cys Ile Ser Thr Leu Tyr |  |     |  |     |
|                 | 395   |  | 400 |  | 405 |
| Arg Asp Pro Ala | Trp Ser Gln Val Ala Leu Gln Gly Pro Ile Glu |  |     |  |     |
|                 | 410   |  | 415 |  | 420 |
| Arg Ala Gln Val | His Val Cys Ser Ser Lys Met Gly Ala Leu Ser |  |     |  |     |
|                 | 425   |  | 430 |  | 435 |
| Thr Glu Arg Leu | Gln Tyr Tyr Thr Gln Glu Leu Gly Val Arg Glu |  |     |  |     |
|                 | 440   |  | 445 |  | 450 |
| Arg Ser Gly His | Ser Val Ser Leu Ile Asp Leu Trp Gly Leu Leu |  |     |  |     |
|                 | 455   |  | 460 |  | 465 |
| Val Glu Tyr Leu | Leu Tyr Gln Glu Glu Asn Pro Ala Lys Leu Ser |  |     |  |     |
|                 | 470   |  | 475 |  | 480 |
| Asp Gln Gln Lys | Ala Val Arg Gln Gly Gln Asn Pro Tyr Pro Ile |  |     |  |     |
|                 | 485   |  | 490 |  | 495 |

|                 |                     |                     |     |
|-----------------|---------------------|---------------------|-----|
| Tyr Thr Ser Val | Asn Val Arg Thr Asn | Leu Ser Gly Glu Asp | Phe |
| 500             |                     | 505                 | 510 |
| Ala Trp Cys Glu | Phe Thr Pro Tyr Lys | Val Gly Phe Pro Lys | Tyr |
| 515             |                     | 520                 | 525 |
| Gly Ala Tyr Val | Pro Thr Glu Leu Phe | Gly Ser Glu Leu Phe | Met |
| 530             |                     | 535                 | 540 |
| Gly Arg Leu Leu | Gln Leu Gln Pro Glu | Pro Arg Ile Cys Tyr | Leu |
| 545             |                     | 550                 | 555 |
| Gln Gly Met Trp | Gly Ser Ala Phe Ala | Thr Ser Leu Asp Glu | Ile |
| 560             |                     | 565                 | 570 |
| Phe Leu Lys Thr | Ala Gly Ser Gly Leu | Ser Phe Leu Glu Trp | Tyr |
| 575             |                     | 580                 | 585 |
| Arg Gly Ser Val | Asn Ile Thr Asp Asp | Cys Gln Lys Pro Gln | Leu |
| 590             |                     | 595                 | 600 |
| His Asn Pro Ser | Arg Leu Arg Thr Arg | Leu Leu Thr Pro Gln | Gly |
| 605             |                     | 610                 | 615 |
| Pro Phe Ser Gln | Ala Val Leu Asp Ile | Phe Thr Ser Arg Phe | Thr |
| 620             |                     | 625                 | 630 |
| Ser Ala Gln Ser | Phe Asn Phe Thr Arg | Gly Leu Cys Leu His | Lys |
| 635             |                     | 640                 | 645 |
| Asp Tyr Val Ala | Gly Arg Glu Phe Val | Ala Trp Lys Asp Thr | His |
| 650             |                     | 655                 | 660 |
| Pro Asp Ala Phe | Pro Asn Gln Leu Thr | Pro Met Arg Asp Cys | Leu |
| 665             |                     | 670                 | 675 |
| Tyr Leu Val Asp | Gly Gly Phe Ala Ile | Asn Ser Pro Phe Pro | Leu |
| 680             |                     | 685                 | 690 |
| Ala Leu Leu Pro | Gln Arg Ala Val Asp | Leu Ile Leu Ser Phe | Asp |
| 695             |                     | 700                 | 705 |
| Tyr Ser Leu Glu | Ala Pro Phe Glu Val | Leu Lys Met Thr Glu | Lys |
| 710             |                     | 715                 | 720 |
| Tyr Cys Leu Asp | Arg Gly Ile Pro Phe | Pro Ser Ile Glu Val | Gly |
| 725             |                     | 730                 | 735 |
| Pro Glu Asp Val | Glu Glu Ala Arg Glu | Cys Tyr Leu Phe Ala | Lys |
| 740             |                     | 745                 | 750 |
| Ala Glu Asp Pro | Arg Ser Pro Ile Val | Leu His Phe Pro Leu | Val |
| 755             |                     | 760                 | 765 |
| Asn Arg Thr Phe | Arg Thr His Leu Ala | Pro Gly Val Glu Arg | Gln |
| 770             |                     | 775                 | 780 |
| Thr Ala Glu Glu | Lys Ala Phe Gly Asp | Phe Val Ile Asn Arg | Pro |
| 785             |                     | 790                 | 795 |
| Asp Thr Pro Tyr | Gly Met Met Asn Phe | Thr Tyr Glu Pro Gln | Asp |
| 800             |                     | 805                 | 810 |
| Phe Tyr Arg Leu | Val Ala Leu Ser Arg | Tyr Asn Val Leu Asn | Asn |
| 815             |                     | 820                 | 825 |
| Val Glu Thr Leu | Lys Cys Ala Leu Gln | Leu Ala Leu Asp Arg | His |
| 830             |                     | 835                 | 840 |
| Gln Ala Arg Glu | Arg Ala Gly Ala     |                     |     |
| 845             |                     |                     |     |

&lt;210&gt; 4

&lt;211&gt; 474

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2970737CD1

&lt;400&gt; 4

|                     |                 |                     |     |
|---------------------|-----------------|---------------------|-----|
| Met Ser Gln Gln Lys | Cys Ile Val Ile | Phe Ala Leu Val Cys | Cys |
| 1                   | 5               | 10                  | 15  |
| Phe Ala Ile Leu Val | Ala Leu Ile Phe | Ser Ala Val Asp Ile | Met |
| 20                  | 25              | 30                  |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Glu | Asp | Glu | Asp | Gly | Leu | Ser | Glu | Lys | Asn | Cys | Gln | Asn | Lys |
|     |     |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |
| Cys | Arg | Ile | Ala | Leu | Val | Glu | Asn | Ile | Pro | Glu | Gly | Leu | Asn | Tyr |
|     |     |     |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |
| Ser | Glu | Asn | Ala | Pro | Phe | His | Leu | Ser | Leu | Phe | Gln | Gly | Trp | Met |
|     |     |     |     | 65  |     |     |     |     | 70  |     |     |     |     | 75  |
| Asn | Leu | Leu | Asn | Met | Ala | Lys | Lys | Ser | Val | Asp | Ile | Val | Ser | Ser |
|     |     |     |     | 80  |     |     |     |     | 85  |     |     |     |     | 90  |
| His | Trp | Asp | Leu | Asn | His | Thr | His | Pro | Ser | Ala | Cys | Gln | Gly | Gln |
|     |     |     |     | 95  |     |     |     |     | 100 |     |     |     |     | 105 |
| Arg | Leu | Phe | Glu | Lys | Leu | Leu | Gln | Leu | Thr | Ser | Gln | Asn | Ile | Glu |
|     |     |     |     | 110 |     |     |     |     | 115 |     |     |     |     | 120 |
| Ile | Lys | Leu | Val | Ser | Asp | Val | Thr | Ala | Asp | Ser | Lys | Val | Leu | Glu |
|     |     |     |     | 125 |     |     |     |     | 130 |     |     |     |     | 135 |
| Ala | Leu | Lys | Leu | Lys | Gly | Ala | Glu | Val | Thr | Tyr | Met | Asn | Met | Thr |
|     |     |     |     | 140 |     |     |     |     | 145 |     |     |     |     | 150 |
| Ala | Tyr | Asn | Lys | Gly | Arg | Leu | Gln | Ser | Ser | Phe | Trp | Ile | Val | Asp |
|     |     |     |     | 155 |     |     |     |     | 160 |     |     |     |     | 165 |
| Lys | Gln | His | Val | Tyr | Ile | Gly | Ser | Ala | Gly | Leu | Asp | Trp | Gln | Ser |
|     |     |     |     | 170 |     |     |     |     | 175 |     |     |     |     | 180 |
| Leu | Gly | Gln | Met | Lys | Glu | Leu | Gly | Val | Ile | Phe | Tyr | Asn | Cys | Ser |
|     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |     | 195 |
| Cys | Leu | Val | Leu | Asp | Leu | Gln | Arg | Ile | Phe | Ala | Leu | Tyr | Ser | Ser |
|     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     | 210 |
| Leu | Lys | Phe | Lys | Ser | Arg | Val | Pro | Gln | Thr | Trp | Ser | Lys | Arg | Leu |
|     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     | 225 |
| Tyr | Gly | Val | Tyr | Asp | Asn | Glu | Lys | Lys | Leu | Gln | Leu | Gln | Leu | Asn |
|     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Glu | Thr | Lys | Ser | Gln | Ala | Phe | Val | Ser | Asn | Ser | Pro | Lys | Leu | Phe |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |
| Cys | Pro | Lys | Asn | Arg | Ser | Phe | Asp | Ile | Asp | Ala | Ile | Tyr | Ser | Val |
|     |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |
| Ile | Asp | Asp | Ala | Lys | Gln | Tyr | Val | Tyr | Ile | Ala | Val | Met | Asp | Tyr |
|     |     |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |
| Leu | Pro | Ile | Ser | Ser | Thr | Ser | Thr | Lys | Arg | Thr | Tyr | Trp | Pro | Asp |
|     |     |     |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |
| Leu | Asp | Ala | Lys | Ile | Arg | Glu | Ala | Leu | Val | Leu | Arg | Ser | Val | Arg |
|     |     |     |     | 305 |     |     |     |     | 310 |     |     |     |     | 315 |
| Val | Arg | Leu | Leu | Leu | Ser | Phe | Trp | Lys | Glu | Thr | Asp | Pro | Leu | Thr |
|     |     |     |     | 320 |     |     |     |     | 325 |     |     |     |     | 330 |
| Phe | Asn | Phe | Ile | Ser | Ser | Leu | Lys | Ala | Ile | Cys | Thr | Glu | Ile | Ala |
|     |     |     |     | 335 |     |     |     |     | 340 |     |     |     |     | 345 |
| Asn | Cys | Ser | Leu | Lys | Val | Lys | Phe | Phe | Asp | Leu | Glu | Arg | Glu | Asn |
|     |     |     |     | 350 |     |     |     |     | 355 |     |     |     |     | 360 |
| Ala | Cys | Ala | Thr | Lys | Glu | Gln | Lys | Asn | His | Thr | Phe | Pro | Arg | Leu |
|     |     |     |     | 365 |     |     |     |     | 370 |     |     |     |     | 375 |
| Asn | Arg | Asn | Lys | Tyr | Met | Val | Thr | Asp | Gly | Ala | Ala | Tyr | Ile | Gly |
|     |     |     |     | 380 |     |     |     |     | 385 |     |     |     |     | 390 |
| Asn | Phe | Asp | Trp | Val | Gly | Asn | Asp | Phe | Thr | Gln | Asn | Ala | Gly | Thr |
|     |     |     |     | 395 |     |     |     |     | 400 |     |     |     |     | 405 |
| Gly | Leu | Val | Ile | Asn | Gln | Ala | Asp | Val | Arg | Asn | Asn | Arg | Ser | Ile |
|     |     |     |     | 410 |     |     |     |     | 415 |     |     |     |     | 420 |
| Ile | Lys | Gln | Leu | Lys | Asp | Val | Phe | Glu | Arg | Asp | Trp | Tyr | Ser | Pro |
|     |     |     |     | 425 |     |     |     |     | 430 |     |     |     |     | 435 |
| Tyr | Ala | Lys | Thr | Leu | Gln | Pro | Thr | Lys | Gln | Pro | Asn | Cys | Ser | Ser |
|     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |     | 450 |
| Leu | Phe | Lys | Leu | Lys | Pro | Leu | Ser | Asn | Lys | Thr | Ala | Thr | Asp | Asp |
|     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     | 465 |
| Thr | Gly | Gly | Lys | Asp | Pro | Arg | Asn | Val |     |     |     |     |     |     |
|     |     |     |     | 470 |     |     |     |     |     |     |     |     |     |     |

&lt;210&gt; 5

&lt;211&gt; 321



&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5924878CD1

&lt;400&gt; 5

```

Met Ala Ser Ser Gln Gly Lys Asn Glu Leu Lys Leu Ala Asp Trp
 1          5          10          15
Met Ala Thr Leu Pro Glu Ser Met His Ser Ile Pro Leu Thr Asn
          20          25          30
Leu Ala Ile Pro Gly Ser His Asp Ser Phe Ser Phe Tyr Ile Asp
          35          40          45
Glu Ala Ser Pro Val Gly Pro Glu Gln Pro Glu Thr Val Gln Asn
          50          55          60
Phe Val Ser Val Phe Gly Thr Val Ala Lys Lys Leu Met Arg Lys
          65          70          75
Trp Leu Ala Thr Gln Thr Met Asn Phe Thr Gly Gln Leu Gly Ala
          80          85          90
Gly Ile Arg Tyr Phe Asp Leu Arg Ile Ser Thr Lys Pro Arg Asp
          95          100          105
Pro Asp Asn Glu Leu Tyr Phe Ala His Gly Leu Phe Ser Ala Lys
          110          115          120
Val Asn Glu Gly Leu Glu Glu Ile Asn Ala Phe Leu Thr Asp His
          125          130          135
His Lys Glu Val Val Phe Leu Asp Phe Asn His Phe Tyr Gly Met
          140          145          150
Gln Lys Tyr His His Glu Lys Leu Val Gln Met Leu Lys Asp Ile
          155          160          165
Tyr Gly Asn Lys Met Cys Pro Ala Ile Phe Ala Gln Glu Val Ser
          170          175          180
Leu Lys Tyr Leu Trp Glu Lys Asp Tyr Gln Val Leu Val Phe Tyr
          185          190          195
His Ser Pro Val Ala Leu Glu Val Pro Phe Leu Trp Pro Gly Gln
          200          205          210
Met Met Pro Ala Pro Trp Ala Asn Thr Thr Asp Pro Glu Lys Leu
          215          220          225
Ile Gln Phe Leu Gln Ala Ser Ile Thr Glu Arg Arg Lys Lys Gly
          230          235          240
Ser Phe Phe Ile Ser Gln Val Val Leu Thr Pro Lys Ala Ser Thr
          245          250          255
Val Val Lys Gly Val Ala Ser Gly Leu Arg Glu Thr Ile Thr Glu
          260          265          270
Arg Ala Leu Pro Ala Met Met Gln Trp Val Arg Thr Gln Lys Pro
          275          280          285
Gly Glu Ser Gly Ile Asn Ile Val Thr Ala Asp Phe Val Glu Leu
          290          295          300
Gly Asp Phe Ile Ser Thr Val Ile Lys Leu Asn Tyr Val Phe Asp
          305          310          315
Glu Gly Glu Ala Asn Thr
          320

```

&lt;210&gt; 6

&lt;211&gt; 1026

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477093CD1

&lt;400&gt; 6

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Pro | Lys | Cys | Leu | Ser | Phe | His | Phe | Thr | Pro | Thr | Gly | Pro | Asn |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |
| Arg | Lys | Glu | Glu | Gln | Gly | Val | Leu | Arg | Val | Gly | Lys | Ala | Gly | Ala |
|     |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |
| Leu | Gly | Asn | Arg | Phe | Gln | Glu | Glu | Arg | Glu | Thr | Val | Cys | Leu | Ser |
|     |     |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |
| Pro | Leu | His | Cys | Phe | Ser | Gln | Glu | Val | Leu | Val | Glu | Ser | Trp | Leu |
|     |     |     |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |
| Val | Cys | Asp | Asp | Leu | Ser | Phe | Leu | Met | Ala | Lys | Ala | Gln | Val | Val |
|     |     |     |     | 65  |     |     |     |     | 70  |     |     |     |     | 75  |
| Leu | Gly | Leu | Met | Thr | Gly | Ser | Cys | Gln | Met | Pro | Leu | Arg | Val | Met |
|     |     |     |     | 80  |     |     |     |     | 85  |     |     |     |     | 90  |
| Ile | Thr | Pro | Gly | Asp | Gln | Ser | Pro | Glu | Pro | Gln | Gly | Asn | Gln | Leu |
|     |     |     |     | 95  |     |     |     |     | 100 |     |     |     |     | 105 |
| Pro | Ala | Thr | Thr | Pro | Gly | Thr | Arg | Leu | Pro | Pro | Pro | Gly | Cys | Ile |
|     |     |     |     | 110 |     |     |     |     | 115 |     |     |     |     | 120 |
| Phe | Cys | Pro | Glu | Thr | Pro | Gly | Leu | Glu | Glu | Val | Leu | Arg | Leu | Gln |
|     |     |     |     | 125 |     |     |     |     | 130 |     |     |     |     | 135 |
| Val | Leu | Glu | Pro | Cys | Gln | Arg | Phe | Leu | Glu | Met | Leu | Val | Asp | Pro |
|     |     |     |     | 140 |     |     |     |     | 145 |     |     |     |     | 150 |
| Ala | Ser | Trp | Gly | Arg | Gly | Ala | Gly | Ser | Trp | Leu | Ile | Glu | Gly | Trp |
|     |     |     |     | 155 |     |     |     |     | 160 |     |     |     |     | 165 |
| Gln | Ile | His | Ala | Gly | Pro | Ile | Pro | Leu | Gly | Pro | Ser | Leu | Pro | Leu |
|     |     |     |     | 170 |     |     |     |     | 175 |     |     |     |     | 180 |
| Ser | Gln | Arg | Ala | Ser | Ile | Leu | Ala | Arg | Ala | Gly | Pro | Gly | Val | Lys |
|     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |     | 195 |
| Leu | Glu | Gly | Leu | Ala | Trp | Arg | Ala | Cys | His | Leu | Gly | Asp | His | Leu |
|     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     | 210 |
| Ala | Thr | Leu | Thr | Arg | Ala | Pro | Val | Pro | Ala | Asp | Asp | Arg | Gly | Glu |
|     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     | 225 |
| Ala | Ser | Thr | Cys | Trp | Gln | Leu | Thr | Val | Arg | Val | Leu | Glu | Ala | Arg |
|     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Asn | Leu | Arg | Trp | Ala | Asp | Leu | Leu | Ser | Glu | Ala | Asp | Pro | Tyr | Val |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |
| Ile | Leu | Gln | Leu | Ser | Thr | Ala | Pro | Gly | Met | Lys | Phe | Lys | Thr | Lys |
|     |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |
| Thr | Leu | Thr | Asp | Thr | Ser | His | Pro | Val | Trp | Asn | Glu | Ala | Phe | Arg |
|     |     |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |
| Phe | Leu | Ile | Gln | Ser | Gln | Val | Lys | Asn | Val | Leu | Glu | Leu | Ser | Ile |
|     |     |     |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |
| Tyr | Asp | Glu | Asp | Ser | Val | Thr | Glu | Asp | Asp | Ile | Cys | Phe | Lys | Val |
|     |     |     |     | 305 |     |     |     |     | 310 |     |     |     |     | 315 |
| Leu | Tyr | Asp | Ile | Ser | Glu | Val | Leu | Pro | Gly | Lys | Leu | Leu | Arg | Lys |
|     |     |     |     | 320 |     |     |     |     | 325 |     |     |     |     | 330 |
| Thr | Phe | Ser | Gln | Ser | Pro | Gln | Gly | Glu | Glu | Glu | Leu | Asp | Val | Glu |
|     |     |     |     | 335 |     |     |     |     | 340 |     |     |     |     | 345 |
| Phe | Leu | Met | Glu | Glu | Thr | Ser | Asp | Arg | Pro | Glu | Asn | Leu | Ile | Thr |
|     |     |     |     | 350 |     |     |     |     | 355 |     |     |     |     | 360 |
| Asn | Lys | Val | Ile | Val | Ala | Arg | Glu | Leu | Ser | Cys | Leu | Asp | Val | His |
|     |     |     |     | 365 |     |     |     |     | 370 |     |     |     |     | 375 |
| Leu | Asp | Ser | Thr | Gly | Ser | Thr | Ala | Val | Val | Ala | Asp | Gln | Asp | Lys |
|     |     |     |     | 380 |     |     |     |     | 385 |     |     |     |     | 390 |
| Leu | Glu | Leu | Glu | Leu | Val | Leu | Lys | Gly | Ser | Tyr | Glu | Asp | Thr | Gln |
|     |     |     |     | 395 |     |     |     |     | 400 |     |     |     |     | 405 |
| Thr | Ser | Phe | Leu | Gly | Thr | Ala | Ser | Ala | Phe | Arg | Phe | His | Tyr | Met |
|     |     |     |     | 410 |     |     |     |     | 415 |     |     |     |     | 420 |
| Ala | Ala | Leu | Glu | Thr | Glu | Leu | Ser | Gly | Arg | Leu | Arg | Ser | Ser | Arg |
|     |     |     |     | 425 |     |     |     |     | 430 |     |     |     |     | 435 |
| Ser | Asn | Gly | Trp | Asn | Gly | Asp | Asn | Ser | Ala | Gly | Tyr | Leu | Thr | Val |
|     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |     | 450 |
| Pro | Leu | Arg | Pro | Leu | Thr | Ile | Gly | Lys | Glu | Val | Thr | Met | Asp | Val |
|     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     | 465 |
| Pro | Ala | Pro | Asn | Ala | Pro | Gly | Val | Arg | Leu | Gln | Leu | Lys | Ala | Glu |

|                 |                     |                     |     |  |     |
|-----------------|---------------------|---------------------|-----|--|-----|
|                 | 470                 |                     | 475 |  | 480 |
| Gly Cys Pro Glu | Glu Leu Ala Val His | Leu Gly Phe Asn Leu | Cys |  |     |
|                 | 485                 |                     | 490 |  | 495 |
| Ala Glu Glu Gln | Ala Phe Leu Ser Arg | Arg Lys Gln Val Val | Ala |  |     |
|                 | 500                 |                     | 505 |  | 510 |
| Lys Ala Leu Lys | Gln Ala Leu Gln Leu | Asp Arg Asp Leu Gln | Glu |  |     |
|                 | 515                 |                     | 520 |  | 525 |
| Asp Glu Val Pro | Val Val Gly Ile Met | Ala Thr Gly Gly Gly | Ala |  |     |
|                 | 530                 |                     | 535 |  | 540 |
| Arg Ala Met Thr | Ser Leu Tyr Gly His | Leu Leu Ala Leu Gln | Lys |  |     |
|                 | 545                 |                     | 550 |  | 555 |
| Leu Gly Leu Leu | Asp Cys Val Thr Tyr | Phe Ser Gly Ile Ser | Gly |  |     |
|                 | 560                 |                     | 565 |  | 570 |
| Ser Thr Trp Thr | Met Ala His Leu Tyr | Gly Asp Pro Glu Trp | Ser |  |     |
|                 | 575                 |                     | 580 |  | 585 |
| Gln Arg Asp Leu | Glu Gly Pro Ile Arg | Tyr Ala Arg Glu His | Leu |  |     |
|                 | 590                 |                     | 595 |  | 600 |
| Ala Lys Ser Lys | Leu Glu Val Phe Ser | Pro Glu Arg Leu Ala | Ser |  |     |
|                 | 605                 |                     | 610 |  | 615 |
| Tyr Arg Arg Glu | Leu Glu Leu Arg Ala | Glu Gln Gly His Pro | Thr |  |     |
|                 | 620                 |                     | 625 |  | 630 |
| Thr Phe Val Asp | Leu Trp Ala Leu Val | Leu Glu Ser Met Leu | His |  |     |
|                 | 635                 |                     | 640 |  | 645 |
| Gly Gln Val Met | Asp Gln Lys Leu Ser | Gly Gln Arg Ala Ala | Leu |  |     |
|                 | 650                 |                     | 655 |  | 660 |
| Glu Arg Gly Gln | Asn Pro Leu Pro Leu | Tyr Leu Ser Leu Asn | Val |  |     |
|                 | 665                 |                     | 670 |  | 675 |
| Lys Glu Asn Asn | Leu Glu Thr Leu Asp | Phe Lys Glu Trp Val | Glu |  |     |
|                 | 680                 |                     | 685 |  | 690 |
| Phe Ser Pro Tyr | Glu Val Gly Phe Leu | Lys Tyr Gly Ala Phe | Val |  |     |
|                 | 695                 |                     | 700 |  | 705 |
| Pro Pro Glu Leu | Phe Gly Ser Glu Phe | Phe Met Gly Arg Leu | Met |  |     |
|                 | 710                 |                     | 715 |  | 720 |
| Arg Arg Ile Pro | Glu Pro Arg Ile Cys | Phe Leu Glu Ala Ile | Trp |  |     |
|                 | 725                 |                     | 730 |  | 735 |
| Ser Asn Ile Phe | Ser Leu Asn Leu Leu | Asp Ala Trp Tyr Asp | Leu |  |     |
|                 | 740                 |                     | 745 |  | 750 |
| Thr Ser Ser Gly | Glu Ser Trp Lys Gln | His Ile Lys Asp Lys | Thr |  |     |
|                 | 755                 |                     | 760 |  | 765 |
| Arg Ser Leu Glu | Lys Glu Pro Leu Thr | Thr Ser Gly Thr Ser | Ser |  |     |
|                 | 770                 |                     | 775 |  | 780 |
| Arg Leu Glu Ala | Ser Trp Leu Gln Pro | Gly Thr Ala Leu Ala | Gln |  |     |
|                 | 785                 |                     | 790 |  | 795 |
| Ala Phe Lys Gly | Phe Leu Thr Gly Arg | Pro Leu His Gln Arg | Ser |  |     |
|                 | 800                 |                     | 805 |  | 810 |
| Pro Asn Phe Leu | Gln Gly Leu Gln Leu | His Gln Asp Tyr Cys | Ser |  |     |
|                 | 815                 |                     | 820 |  | 825 |
| His Lys Asp Phe | Ser Thr Trp Ala Asp | Tyr Gln Leu Asp Ser | Met |  |     |
|                 | 830                 |                     | 835 |  | 840 |
| Pro Ser Gln Leu | Thr Pro Lys Glu Pro | Arg Leu Cys Leu Val | Asp |  |     |
|                 | 845                 |                     | 850 |  | 855 |
| Ala Ala Tyr Phe | Ile Asn Thr Ser Ser | Pro Ser Met Phe Arg | Pro |  |     |
|                 | 860                 |                     | 865 |  | 870 |
| Gly Arg Arg Leu | Asp Leu Ile Leu Ser | Phe Asp Tyr Ser Leu | Ser |  |     |
|                 | 875                 |                     | 880 |  | 885 |
| Ala Pro Phe Glu | Ala Leu Gln Gln Thr | Glu Leu Tyr Cys Arg | Ala |  |     |
|                 | 890                 |                     | 895 |  | 900 |
| Arg Gly Leu Pro | Phe Pro Arg Val Glu | Pro Ser Pro Gln Asp | Gln |  |     |
|                 | 905                 |                     | 910 |  | 915 |
| His Gln Pro Arg | Glu Cys His Leu Phe | Ser Asp Pro Ala Cys | Pro |  |     |
|                 | 920                 |                     | 925 |  | 930 |
| Glu Ala Pro Ile | Leu Leu His Phe Pro | Leu Val Asn Ala Ser | Phe |  |     |
|                 | 935                 |                     | 940 |  | 945 |

```

Lys Asp His Ser Ala Pro Gly Val Gln Arg Ser Pro Ala Glu Leu
      950      955      960
Gln Gly Gly Gln Val Asp Leu Thr Gly Ala Thr Cys Pro Tyr Thr
      965      970      975
Leu Ser Asn Met Thr Tyr Lys Glu Glu Asp Phe Glu Arg Leu Leu
      980      985      990
Arg Leu Ser Asp Tyr Asn Val Gln Thr Ser Gln Gly Ala Ile Leu
      995      1000      1005
Gln Ala Leu Arg Thr Ala Leu Lys His Arg Thr Leu Glu Ala Arg
      1010      1015      1020
Pro Pro Arg Ala Gln Thr
      1025

```

&lt;210&gt; 7

&lt;211&gt; 379

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2194717CD1

&lt;400&gt; 7

```

Met Ile Ile Tyr Thr Ser Gly Thr Thr Gly Arg Pro Lys Gly Val
  1      5      10      15
Leu Ser Thr His Gln Asn Ile Arg Ala Val Thr Gly Leu Val
      20      25      30
His Lys Trp Ala Trp Thr Lys Asp Asp Val Ile Leu His Val Leu
      35      40      45
Pro Leu His His Val His Gly Val Val Asn Ala Leu Leu Cys Pro
      50      55      60
Leu Trp Val Gly Ala Thr Cys Val Met Met Pro Glu Phe Ser Pro
      65      70      75
Gln Gln Val Trp Glu Lys Phe Leu Ser Ser Glu Thr Pro Arg Ile
      80      85      90
Asn Val Phe Met Ala Val Pro Thr Ile Tyr Thr Lys Leu Met Glu
      95      100      105
Tyr Tyr Asp Arg His Phe Thr Gln Pro His Ala Gln Asp Phe Leu
      110      115      120
Arg Ala Val Cys Glu Glu Lys Ile Arg Leu Met Val Ser Gly Ser
      125      130      135
Ala Ala Leu Pro Leu Pro Val Leu Glu Lys Trp Lys Asn Ile Thr
      140      145      150
Gly His Thr Leu Leu Glu Arg Tyr Gly Met Thr Glu Ile Gly Met
      155      160      165
Ala Leu Ser Gly Pro Leu Thr Thr Ala Met Arg Leu Pro Gly Ser
      170      175      180
Val Gly Thr Pro Leu Pro Gly Val Gln Val Arg Ile Val Ser Glu
      185      190      195
Asn Pro Gln Arg Glu Ala Cys Ser Tyr Thr Ile His Ala Glu Gly
      200      205      210
Asp Glu Arg Gly Thr Lys Val Thr Pro Gly Phe Glu Glu Lys Glu
      215      220      225
Gly Glu Leu Leu Val Arg Gly Pro Ser Val Phe Arg Glu Tyr Trp
      230      235      240
Asn Lys Pro Glu Glu Thr Lys Ser Ala Phe Thr Leu Asp Gly Trp
      245      250      255
Phe Lys Thr Gly Asp Thr Val Val Phe Lys Asp Gly Gln Tyr Trp
      260      265      270
Ile Arg Gly Arg Thr Ser Val Asp Ile Ile Lys Thr Gly Gly Tyr
      275      280      285
Lys Val Ser Ala Leu Glu Val Glu Trp His Leu Leu Ala His Pro
      290      295      300

```

```

Ser Ile Thr Asp Val Ala Val Ile Gly Val Pro Asp Met Thr Trp
      305                      310                      315
Gly Gln Arg Val Thr Ala Val Val Thr Leu Arg Glu Gly His Ser
      320                      325                      330
Leu Ser His Arg Glu Leu Lys Glu Trp Ala Arg Asn Val Leu Ala
      335                      340                      345
Pro Tyr Ala Val Pro Ser Glu Leu Val Leu Val Glu Glu Ile Pro
      350                      355                      360
Arg Asn Gln Met Gly Lys Ile Asp Lys Lys Ala Leu Ile Arg His
      365                      370                      375
Phe His Pro Ser

```

&lt;210&gt; 8

&lt;211&gt; 362

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473574CD1

&lt;400&gt; 8

```

Met Leu Cys Ala Leu Leu Leu Pro Ser Leu Leu Gly Ala Thr
  1      5      10      15
Arg Ala Ser Pro Thr Ser Gly Pro Gln Glu Cys Ala Lys Gly Ser
      20      25      30
Thr Val Trp Cys Gln Asp Leu Gln Thr Ala Ala Arg Cys Gly Ala
      35      40      45
Val Gly Tyr Cys Gln Gly Ala Val Trp Asn Lys Pro Thr Ala Lys
      50      55      60
Ser Leu Pro Cys Asp Val Cys Gln Asp Ile Ala Ala Ala Ala Gly
      65      70      75
Asn Gly Leu Asn Pro Asp Ala Thr Glu Ser Asp Ile Leu Ala Leu
      80      85      90
Val Met Lys Thr Cys Glu Trp Leu Pro Ser Gln Glu Ser Ser Ala
      95     100     105
Gly Cys Lys Trp Met Val Asp Ala His Ser Ser Ala Ile Leu Ser
     110     115     120
Met Leu Arg Gly Ala Pro Asp Ser Ala Pro Ala Gln Val Cys Thr
     125     130     135
Ala Leu Ser Leu Cys Glu Pro Leu Gln Arg His Leu Ala Thr Leu
     140     145     150
Arg Pro Leu Ser Lys Glu Asp Thr Phe Glu Ala Val Ala Pro Phe
     155     160     165
Met Ala Asn Gly Pro Leu Thr Phe His Pro Arg Gln Ala Pro Glu
     170     175     180
Gly Ala Leu Cys Gln Asp Cys Val Arg Gln Leu Val Ala Lys Ile
     185     190     195
Thr Pro Glu Lys Val Cys Lys Phe Ile Arg Leu Cys Gly Asn Arg
     200     205     210
Arg Arg Ala Arg Ala Val His Asp Ala Tyr Ala Ile Val Pro Ser
     215     220     225
Pro Glu Trp Asp Ala Glu Asn Gln Gly Ser Phe Cys Asn Gly Cys
     230     235     240
Lys Arg Leu Leu Thr Val Ser Ser His Asn Leu Glu Ser Lys Ser
     245     250     255
Thr Lys Arg Asp Ile Leu Val Ala Phe Lys Gly Gly Cys Ser Ile
     260     265     270
Leu Pro Leu Pro Tyr Met Ile Gln Cys Lys His Phe Val Thr Gln
     275     280     285
Tyr Glu Pro Val Leu Ile Glu Ser Leu Lys Asp Met Met Asp Pro
     290     295     300

```

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Ala | Val | Cys | Lys | Lys | Val | Gly | Ala | Cys | His | Gly | Pro | Arg | Thr |
|     |     |     |     | 305 |     |     |     |     | 310 |     |     |     |     | 315 |
| Pro | Leu | Leu | Gly | Thr | Asp | Gln | Cys | Ala | Leu | Gly | Pro | Ser | Phe | Trp |
|     |     |     |     | 320 |     |     |     |     | 325 |     |     |     |     | 330 |
| Cys | Arg | Ser | Gln | Glu | Ala | Ala | Lys | Leu | Cys | Asn | Ala | Val | Gln | His |
|     |     |     |     | 335 |     |     |     |     | 340 |     |     |     |     | 345 |
| Cys | Gln | Lys | His | Val | Trp | Lys | Glu | Met | His | Leu | His | Ala | Gly | Glu |
|     |     |     |     | 350 |     |     |     |     | 355 |     |     |     |     | 360 |
| His | Ala |     |     |     |     |     |     |     |     |     |     |     |     |     |

&lt;210&gt; 9

&lt;211&gt; 3594

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2181310CB1

&lt;400&gt; 9

```

gatgggctgg actggagaga gctcacacct ctccccttct tactgcttcc ctccggctat 60
aacttgccag tcacagcagc cagctgctgt agaagagggg aggaacaag ccagtgcag 120
gggagcaaaa gagaaaagga gccaggtctg gcttctgat cccacagcat cgcagagctc 180
gggaggcaca gctcacagac acaggaaca caggactgct attctgctct cctgcccacg 240
gtgatctggt gccagctggt ggaacagtgg gtgatggcgt ccctgctgca agaccagctg 300
accactgac aggacttgct gctgatgcag gaaggcatgc cgatgcgcaa ggtgaggctc 360
aaaagctgga agaagctaag atacttcaga cttcagaatg acggcatgac agtctggcat 420
gcacggcagg ccaggggcag tgccaagccc agcttctcaa tctctgatgt ggagacaata 480
cgtaatggcc atgattccga gttgctgctg agcctggcag aggagctccc cctggagcag 540
ggcttcacca ttgtcttcca tggccgcgcg tccaacctgg acctgatggc caacagtgtt 600
gaggaggccc agatatggat gcgagggtc cagctgttgg tggatcttgt caccagcatg 660
gaccatcagg agcgctgga ccagtggctg agcgattggt ttcaacgtgg agacaaaaat 720
caggatggta agatgagttt ccaagaagtt cagcggttat tgcacctaat gaatgtgga 780
atggaccaag aatatgcctt cagtcttttt caggagcttc ggaggaagat cctggtgaag 840
gggaagaagt taacacttga ggaagacctg gaatatgagg aagaggaagc agaacctgag 900
ttggaagagt cagaattggc gctggagtcc cagtttgaga ctgagcctgg caaactgcgg 960
catgtgctga gtatggatgg ettcctcagc tacctctgct ctaaggatgg agacatcttc 1020
aaccagcct gcctcccat ctatcaggat atgactcaac ccctgaacca ctacttcac 1080
tgctcttctc ataacaccta cctagtgggg gaccagcttt gtggccagag cagcgtcgag 1140
ggatatatac ggagggccct gaagcggggg tgccgctgcg tggaggtgga tgtatgggat 1200
ggacctagcg gggaacctgt cgtttaccac ggacacaccc tgacctccc catcctgttc 1260
aaagatgtcg tggccacagt agcacagtat gccttcaga catcagacta cccagtcac 1320
ttgtccctgg agaccactg cagctgggag cagcagcaga ccatggcccc tcatctgact 1380
gagatcctgg gggagcagct gctgagcacc accttggatg ggggtgctgcc cactcagctg 1440
ccctcgctg aggagcttcg gaggaagatc ctggtgaagg ggaagaagtt aacacttgag 1500
gaagacctgg aatatgagga agaggaagca gaacctgagt tggaagagtc agaattggcg 1560
ctggagtcac agtttgagac tgagcctgag cccaggagc agaaccttca gaataaggac 1620
aaaaagaaga aatccaagcc catcttgtgt ccagccctc cttccctggt tatctacttg 1680
aagtctgtct cattccgcag cttcacacat tcaaaggagc actaccactt ctacgagata 1740
tcatctttct ctgaaaccaa ggccaagcgc ctcatcaagg aggctggcaa tgagtttgtg 1800
cagcacaata cttggcagtt aagccgtgtg tatcccagcg gcctgaggac agactcttcc 1860
aactacaacc cccaggaaact ctggaatgca ggctgccaga tgggtggccat gaatatgcag 1920
actgcagggc ttgaaatgga catctgtgat gggcatttcc gccagaatgg cggctgtggc 1980
tatgtgctga agccagactt cctgcgtgat atccagagtt ctttccacc tgagaagccc 2040
atcagccctt tcaaagccca gactctctta atccaggtga tcagcgggtca gcaactcccc 2100
aaagtggaca agaccaaaga ggggtccatt gtggatccac tgggtgaaagt gcagatcttt 2160
ggcgttcgtc tagacacagc acggcaggag accaactatg tggagaacaa tggttttaat 2220
ccatactggg ggcagacact atgtttccgg gtgctggtgc ctgaacttgc catgctgcgt 2280
tttgggtaaa tggattatga ctggaaatcc cgaaatgact ttattgggtc gtacaccctg 2340
ccttggacct gcatgcaaca aggttaccgc cacattcacc tgctgtccaa agatggcatc 2400
agcctccgcc cagcttccat ctttgtgtat atctgcatcc aggaaggcct ggagggggat 2460
gagtcctgag gtgggcattt cacgggaagg gttggtgtgc tggctttaga cggggagaaa 2520

```

```

catctggaag gatgctcgag agaacaaatg gaggtggtga aaatcaagct ttggattgtg 2580
cattcctagg cacaaaatta cctcattctt cctaacaagc aatctgggac ctgattttcc 2640
accttttttc tcttttcttc ccttcctttg ttttcataag cctttggtat ctttcctgcc 2700
cttttccttt gtgtactcta tactggagtt cctttcttcc tcttgctgta ggctcaatcc 2760
cataccgaca tctacaacta atctttccca tcaactctgt gtgaaggcag gttgcaacta 2820
gaaattcaga ggggcttgga atagagaaac ctaaagaagc atcatcccct ccatcccaa 2880
cttctcaaaa gcccaaagcc aagggaagga taaatcaagg ctcaaggctt cccagcaaa 2940
gattagggaa agagacttga cccaggaact gtactacgac tcttaagaga aactgcaca 3000
gcactcaaa gctccccactg gactgcttcc tccttagccc cactggtata aatacatctc 3060
tctccaattt ggcttcaata tggctctgtca ttgttgggca agaaggggag gtacaagggt 3120
tgtggggaca tctgggtagt caggtgagac aaggaaaagg tagagaaaga ggttcccagg 3180
agacctcttg catgtgctac ataggaagga cacagagtat ggcttttaag aatcagggca 3240
aaagcagaca aaaggttatg tgggtcccagc ctccctgaag agtctggctg gaaccatcct 3300
agtttatgtc tgttgctcca gcataattat taatagtacc ctctttaagt tttcccagt 3360
tgtctcagtt tagatgatca attatatagt aaggcactac agaattatac tgtgaagcag 3420
ttatagaatt ctgaattgga gattttcttc tctagtcac tgcttttoca tagtatcctt 3480
gtgctttaac accttagtgt agctgctgag ccagtctgag gattgatgat gagctaaagt 3540
tgcttctaag aggaggctta gtgggtagca atcctccctc tatagataca ggta 3594

```

&lt;210&gt; 10

&lt;211&gt; 2243

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2965233CB1

&lt;400&gt; 10

```

ggccctgagg ctgaggggct gagtgetcat tccagccgcc tcggggaacc cgggctggga 60
gaccccatgc ctgggggtga gcctggagcc agggcagtg ggtgagaggc tccggagaga 120
gggctgggca ccaccaggct tgggtgtgtg atgcgctgct ggcccaggct acaccccgac 180
aagggaacac gggggccccc ggagcagaga gacctcagag cagcctcctc ctgctcctg 240
tggacggccg gcccagctg gtgateccag ccagtcacag ctttcagttg ctgccccac 300
cgacagtcct cagtccttcc atgatggctc ccccgacagc cggcccccct cctggcccag 360
ctcttccgcc tgaggaccce gggccggatc cggagagcag gtggcttttc ttgagcgcca 420
acattctgcc cgtggtggag cgggtgcattg gtgccatgca agaggggatg cagatgggtga 480
agctgcgtgg cggctccaag ggcttggctc gcttctacta cctggacgag caccgctcct 540
gcacccgctg gaggccctca cgcaagaacg agaaggccaa gatctccatc gactccatcc 600
aggaggtgag tgaggggccc cagtcggagg tcttccagcg ctaccctgac ggcagcttcg 660
accccaactg ctgcttcagc atctaccacg gcagccaccg cgagtcgctg gacctggtct 720
ccaccagcag cgaggtggcg cgcacctggg tctactggct gcgctacctc atggccgca 780
tcagcgacga ggacagcctg gctcgcgcgc agcgaccag ggaccagtgg ctgaagcaga 840
cgtttgacga ggccgacaag aacggggatg gcagcctgag cattggcgag gtcttgacg 900
tgctgcacaa gctcaacgtg aacctgcccc ggcagagggg gaagcagatg ttcagggaag 960
cggacacgga tgaccaccaa ggacgctgg gtttgaga gttctgtgcc ttctacaaga 1020
tgatgtccac ccgcccggac ctctacctgc tcatgctgac ctacagcaac cacaaggacc 1080
acctggatgc cgccagcctg cagcgcttcc tgcaggtgga gcagaagatg gcggtgtgta 1140
ccctcgagag ctgccaggac atcatcgagc agtttgagcc atgcccagaa aacaagagta 1200
aggggctgct gggcattgat ggcttcacca actacaccag gagccctgct ggtgacatct 1260
tcaacctga gcaccacat gtgcaccagg acatgacgca gccgctgagc cactacttca 1320
tcacctcgtc ccacaacacc tacctcgtgg gtgaccagct catgtcccag tcacgggtgg 1380
acatgtatgc ttgggtcctg caggctggct gccgctgct ggaggtggac tgctgggatg 1440
ggcccgacgg ggagcccatt gtgcaccatg gctacactct gacttccaag atcctcttca 1500
aagacgtcat tgaaccatc aacaaatag ctttcatcaa gaatgagtac ccagtgatcc 1560
tgtccatcga aaaccactgc agtgtcatcc agcagaagaa aatggcccag tatctgactg 1620
acatccttgg ggacaagctg gacctgtcat cagtgaagcag tgaagatgcc accacactcc 1680
cctctccaca gatgetcaag ggcaagatcc tcgtgaaggg gaagaagctc ccagccaaca 1740
tcagcgagga tgcggaggaa ggcgaggtgt ctgatgagga cagtgcctgat gagattgacg 1800
atgactgcaa gctcctcaat ggggatgcat ccaccaatcg aaagcgtgta gaaaacactg 1860
ctaagaggaa actggattcc ctcatcaaag agtcgaagat tcgggactgt gaggacccca 1920
acaacttctc cgtctccaca ctgtcccat ctggaagct cggaacgcaag agcaaggctg 1980
aagaggacgt ggagtctggg gaggatgccg gggccagcag acgcaatggc cgctcgtcg 2040

```

```

tgggaagctt ctccaggcgc aagaagaagg gcagcaagct gaagaaggcg gccagcgtgg 2100
aggagggaga tgagggtcag gactccccgg gaggccagag ccgaggggcg acccggcaga 2160
agagttctgt gccttctaca agatgatgtc caccgcgcgg gacctctacc tgctcatgct 2220
gacctacagc aaccataaaa ggg                                     2243

```

```

<210> 11
<211> 2547
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 1281946CB1

```

```

<400> 11
atgcttttggg cactctggcc aagggtggctg gcagacaaga tgctgcccct cctgggggca 60
gtgctgtcttc agaagagaga gaagagggtc cctctgtgga ggcactggcg gcgggaaacc 120
taccataact atgacctcca ggtgaagggtg ctgagggcca caaacatccg gggcacagac 180
ctgctgtcca aagccgactg ctatgtgcaa ctgtggctgc ccacggcgtc cccaagccct 240
gccagacta ggatagtggc caactgcagt gaccccgagt ggaatgagac cttccactac 300
cagatccatg gtgctgtgaa gaacgtctcg gagctcacc tttatgacaa ggacatcctg 360
ggcagcgacc agctctccct gctcctgttt gacctgagaa gcctcaagtg tggccaacct 420
cacaacaca ccttccact caaccaccag gattcacaag agctgcaggt ggaatttgtt 480
ctggagaaga gccagggtgcc tgcatctgaa gtcacacca acgggggttct ggtgggtcac 540
ccctgtctga gaatccaggg cacgctccgg ggagatggga cagccccacg ggaagagtac 600
ggctctaggc agctccagct ggcagtgcct ggagcctacg agaagccaca gctcttgccc 660
ctgcagcctc ccacagagcc aggcctccca cccaccttta cctccacgt gaaccagtg 720
ctgagctcca ggctacacgt ggagctgatg gagctgctgg cagctgtgca gagtggcccc 780
agcgagagat tggagggtca gaccagcaag ctgggagagg ggggcatcct gctctcctct 840
ctgcccctag gccaggagga acagtgttct gtggccctgg gggaggggga gggagtggct 900
ctgagcatga cgggtggaat gagctccggg gacctagacc tacgccttgg ctttgacctc 960
tctgacgggg agcaggagtt tctggacagg aggaagcagg tcgtgtccaa ggccctggcg 1020
caagtgtctg gattgagtga ggctctggac agtggccagg tgctgtagt ggctgtgtt 1080
ggttccgggg gtggaacccg agccatgtct tctctgtacg gcagcctggc aggggttcag 1140
gagctcggcc ttctagacac tgtgacctac ctgagtgggg tctctgggtc tacctgggtc 1200
atctccacac tctacaggga cccagcctgg tcccagggtg ccttgcaggg ccccatagag 1260
cgtgcccagg ttcacgtctg cagcagtaag atgggagctt tgccacgga gcggctacag 1320
tactacactc aggaactggg ggtccgggag cgcagtggcc acagcgtgtc cctcatcgac 1380
ctctggggcc tccttgttga gtatctctg taccaggagg agaaccctgc caagctgtct 1440
gaccaacaga aggcgggtcc ccagggtcaa aacccttacc ccatctacac cagtgtcaac 1500
gtccgcacca acttgagtgg ggaagatttt gcatggtgcg agttcacgcc ctataaggtt 1560
ggcttcccc agtacggggc ttatgttccc accgagctct tcggctcaga actcttcag 1620
ggacgattgc tgcagctcca gcctgaaccc cggatctgtt acctgcaagg tatgtggggc 1680
agcgcctttg ccaccagcct ggatgagatc ttctaaaga ccgccgggtc gggcctcagc 1740
ttcctggagt ggtacagagg cagtgtgaat atcacagac actgccagaa gcctcagctg 1800
cacaacccct cgagggtgcg aacgaggctc ctcacccac agggggcctt ccccagggt 1860
gtgctggaca tattcacctc ccgcttact tccgccaga gctttaactt caccgggggt 1920
ctctgcttgc acaaggacta tgtggctggc agggagttcg tggcctggaa agacacacac 1980
ccggagcct tccccaaaca gctcaccccc atgcccggact gcctgtacct ggtggacgga 2040
ggctttgcca tcaactctcc gttcccactg gctctgtgc ctcagagagc agtggacctc 2100
attctgtcct ttgactatcc cttggaagcc ccttttgagg tcttgaagat gacagagaag 2160
tactgcctgg accgaggaat ccccttccct agcatcgagg tgggcctga ggacgtggag 2220
gaggcccggt agtgctatct gtttgccaag gctgaggacc cccgctcccc cattgtgctg 2280
cacttcccc tggttaaccg taccttccgc acacacctgg cccaggtgt ggagcgacaa 2340
acagctgagg agaaggcctt tggggacttt gtcacaaaca ggccagacac cccctatggc 2400
atgatgaact tcacctatga gcccaggac ttttatcggc tgggtggcctt cagtcgatac 2460
aacgtcctga acaacgtgga gacctgaag tgcgcctcc agctggctct ggaccggcac 2520
caggctcggg agagggcagg ggcctga                                     2547

```

```

<210> 12
<211> 2673
<212> DNA
<213> Homo sapiens

```



<220>  
<221> misc\_feature  
<223> Incyte ID No: 2970737CB1

<400> 12  
acgggggcagg tegtgtgtgc cacatgaaac attctctgcc tctgaaggct ccagaggctg 60  
tggaagcag cacatcagcg acagctcctg gctgctggac tccgcaggga gggaaggaaag 120  
gttggtcgca atgtcccagc agaagtgcac cgtgatcttt gccctggtgt gctgctttgc 180  
cattttggtt gcactgatct ttccagccgt ggacatcatg ggagaggatg aggatggact 240  
ctcagaaaaa aattgccaaa ataaatgtcg aattgccctg gtggaaaata ttcttgaagg 300  
ccttaactat tcagaaaaatg caccatttca cttatcactt ttccaaggct ggatgaattt 360  
actcaacatg gccaaaaagt ctgttgacat agtgtcttcc cattgggatc tcaaccacac 420  
tcattccatca gcatgtcagg gtcaacgtct ttttgaaaag ttgctccagc tgacttcgca 480  
aaatattgaa atcaagctag tgagtgatgt aacagctgat tcaaaggatg tagaagcctt 540  
gaaattaaag ggagccgagg tgacgtacat gaacatgacc gcttacaaca agggccggct 600  
gcagtcctcc ttctggatcg tggacaaaca gcacgtgtat atcggcagtg ccggtttgga 660  
ctggcaatcc ctgggacaga tgaaagaact cgggtgtcatc ttctacaact gcagctgcct 720  
ggctctagat ttacaaagga tatttgcctc atatatgttca ttaaaattca aaagcagagt 780  
gcctcaaacc tgggtccaaa gactctatgg agtctatgac aatgaaaaga aattgcaact 840  
tcagttgaat gaaaccaaact ctcaagcatt tgtatcgaat tctccaaaac tcttttgccc 900  
taaaaacaga agttttgaca tagatgccat ctacagtgtg atagatgatg ccaagcagta 960  
tgtgtacatc gctgtcatgg actacctgcc tatctccagc acaagcacca aaaggactta 1020  
ctggccagac ttggatgcaa aaataagaga agcattagtt ttacgaagcg ttagagttcg 1080  
actcctttta agcttctgga aggaaactga tccccttacg tttaacttta tttcatctct 1140  
taaagcgatt tgcactgaaa tagccaactg cagtttgaaa gttaaatttt ttgatctgga 1200  
aagagagaat gcttgtgcta caaaagaaca aaagaatcac acctttccta gggttaaatcg 1260  
caacaagtac atgggtgacag atggagcagc ttatatgga aattttgatt gggtagggaa 1320  
tgatttcact cagaatgtcg gcacgggcct ttgtatcaac caggcagatg tgaggaacaa 1380  
cagaagcatc attaagcaac ttaaagatgt gtttgaaagg gactggtatt caccgtatgc 1440  
caaaacctta cagccaacca aacagccgaa ctgctcaagc ctgttcaaac tcaaaccctt 1500  
ctccaacaaa actgccacag acgacacagg cggaaaggat ccccggaacg tataacatga 1560  
tgaagaaact gacaggacag ctctgtattt catatttata cataaaggac ttgaggagag 1620  
aaaaaacact ttaatatgtc tcttttttta gggaaaaagc acacttataa aaaatattct 1680  
ctgaacaaca cataatacct atctaacaat atcttagcgt cgtgtacact aaaattaaga 1740  
cttttgtatt tgttacctct gacagagaat gtcattctgg cgtagaagtt agttttttacg 1800  
tttgcataca aatttttaaga ctttcattcc tgctcctttc tagttttaga actttttctg 1860  
ctgacctacc cggtcagttg ttaggaagct tagcatgagg tgggctcttt aacaccattc 1920  
tgagaactgt tgatttagag aggaagggtga ggatttgcct tggagaagag actgataaaa 1980  
ccaagtgttt attcttghtaa atcgtctcca agacctacac agtcagctaa gcatgaatag 2040  
cttacatctc atgccagcct tcaacaccct tgaaaaatat ccttttctgt cttgctcatt 2100  
ttcttcttat accataagtt ttatttggta tcattttctc ttctcacaaa atgttccctat 2160  
tatatagcat tttctgtatt cactcaaca ttacttctaa ctatgataac atctttttgaa 2220  
aacatacaaa tttttcagaa tggataactaa gtggaaaatag cagtctttca ttttcacctt 2280  
tcaactgaat ccttttttcc cctaccgaat ttgtatttaa aaaaaatcac agtcagcaat 2340  
atgtatcctt tctctcaagc tcagaaaaat tatgtagcca cctgctaata tccagcaagc 2400  
tatcaagtct gcattttggg aataaagtag tttttcacat ttttggtcag tttatggat 2460  
tataaaggaa aaatctctat gcgtcagtggt attttaattt ctttagaatc attatgctgt 2520  
taagagtatg ccagcaaacg tttctagagc tatttaatat accttctgct atttaataa 2580  
ccaaatgctt ttcagagaaa tgcaatttaa atactgtgt taacataatt tactaagaag 2640  
cagaaatatt gaattattgt tctataaaaa aaa 2673

<210> 13  
<211> 1422  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 5924878CB1

<400> 13  
acgggctgca agaacaaaag ctccagcgga gtctggcgga ggcactaagg gagcacgggg 60  
tgttttgaat aattattaaa attagcatgt gtctgcgcca tctgtgggt gtggcgca 120

```

gaggagtgtta aactctagcg gggctagagc aaacattgga ttagcggcag cagcctgcc 180
gcctgcccga ggagtgtctg gaccagcgcg ctgcacgccg actggcacga tggcctcgtc 240
tcaggggaaa aacgagctga aattagccga ctggatggca actctgccgg agagcatgca 300
cagcatcccc ctcaccaatt tagccattcc agggctctcat gattccttca gcttctacat 360
tgatgaagcc tctccagtag gtctcgagca gccagaaact gtccagaatt ttgtctctgt 420
gtttggaact gtggccaaaa agctcatgcg gaaatgggta gccactcaga caatgaattt 480
tactggccag ctaggagctg gaattcggtt ttttgatctt cgaatttcca ccaagcccag 540
agaccccgac aatgaactct attttgctca tggtttggtc agtgccaaag tcaatgaagg 600
ccttgaggag atcaatgcat tcctcacaga tcaccataag gaggtagtgt tcttgactt 660
caaccacttt tatgggatgc agaaatatca ccatgaaaaa ctggtccaaa tgctgaaaga 720
catctatgga aataaaatgt gccagcgat ttttgcccag gaagttagtt taaagtacct 780
gtgggagaag gactatcaag tgctggtctt ctaccatagt ccagtggctc tgggaagtgc 840
ctttctcttg cctgggcaga tgatgccagc accctgggcc aacaccacag accccgagaa 900
actgatccag tttcttcaag catccatcac tgagagaaga aagaagggat cgttttttat 960
atctcaggtg gtgctgaccc ccaaagctag cactgtgggc aaaggggtgg caagtggcct 1020
cagagaaaca atcacagaaa gagctcttcc tgccatgatg cagtgggtcc gcacgcagaa 1080
gccaggagag atgggcatac atattgtcac tgccgatttt gtagaacttg gtgactttat 1140
cagcactgtc ataaagctca actatgtctt tgatgaagga gaagccaaca cttgatagca 1200
ctacttgagg tttccatgaa taagatggag aaagctcatt gtattagggc atactatctg 1260
taacactctg atcttcctat tccatctgag tctactgaag ggtatagggc tgttatgtgg 1320
gtgggatata tggggaaaac tgtttcttca tgtgattaca atcatgctct tcatcacgta 1380
taaatatttc atccccgcgg tcgagtaaca acatcatcag ct 1422

```

&lt;210&gt; 14

&lt;211&gt; 3197

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7477093CB1

&lt;400&gt; 14

```

atgcctaagt gcctctcctt ccacttcacc ccaacaggcc ctaaccgaaa ggaggagcaa 60
ggagtgtctc gtgtgggcaa agcgggtgct cttggaaatc gattccagga agagagagaa 120
acagtctgcc tctctccact ccactgtttc tcgcaagagg tccttgtgga gtcttggctg 180
gtctgcgatg acctctcttt cctgatggca aaagctcaag tagtccttgg tttgatgact 240
ggctcctgtc agatgcctct cggggtcatg ataaccacag gggaccagag ccctgagcct 300
caggggaacc agcttctctg caccactcct gggactcgtc tgcttcccc tggtctgcatc 360
ttctgtctct aaactccagg tctggaagag gtgctecgcc tgcaggctct ggagccctgc 420
cagagggtttc tggagatgct ggttgacca gcctcttggg gaagaggggc aggatcgtgg 480
ctcattgaag gctggcagat tcatgtggt cctatccac tgggaccgtc gctgcccactc 540
agccagagag ccagcatctt ggcaagggtc gggcctggag tgaagctgga agggctagca 600
tggagagcct gtcacctggg ggaccacctg gccacctta ccagggcgcc agtgccagca 660
gacgacaggg gggaggcctc tacctgctgg cagctcacag tgagggtcct ggaggcgcg 720
aacctgcgct gggctgacct gttgagttag gccgacctt acgtgatect acagctgtcg 780
accgacactg gaatgaagtt taagaccaag acgctcaccg acaccagtca tctgtgttg 840
aatgaggcct tccgtttcct tatccaaagt caggtaaga atgttctgga gcttagcatc 900
tatgatgagg actcagtcac ggaggatgac atctgcttca aggttctcta tgacatctca 960
gaagtccctc ctggcaagct gctccgaaa accttctccc agagtcccca gggagaggag 1020
gagctggatg tggagtccct gatggaagaa acgtcagatc gccagaaaa cctcatcacc 1080
aacaagtgca ttgtggcccg agagctgtca tgcttgatg tgcacttgga cagcacaggg 1140
agcaccgctg tggttgcaga tcaggacaag ctggagctgg agctgggtgct gaaggggtcc 1200
tatgaggaca cacagacatc cttcctgggc acagcctctg ccttcgctt ccactacatg 1260
gcagccctag agacagagct gagcgggcgc ctgaggagct ccagaagcaa tggctggaat 1320
ggggacaact cagctgggta cctcactgtg cccttgaggc ccttgaccat tgggaaggag 1380
gtgactatgg atgttctctg tccaaatgcc ccaggagtga ggctgcagct caaggcagag 1440
ggctgccctg aggagctggc cgtgcacctg ggcttcaatc tctgtgcaga ggagcaggcc 1500
ttcctgagca ggaggaagca ggtggtggcc aagggcctga agcaggccct gcagctggac 1560
agagacctgc aggaggtga ggtaccctca tggggcatca tggccacagg aggaggtgcc 1620
cgggccatga cctcactcta cggccacctt ttggccttgc agaagctggg cctcctagac 1680
tgtgtgacct acttcagtgg catctctggc tctacgtgga caatggccca cctgtacggg 1740
gacctgaggt ggtcgcagag ggacctggag ggacctatca gatacgcccc ggagcacctg 1800

```

```

gccaaagagca agctggaggt cttttcccca gagcgcctgg cgagctaccg ccgggagctg 1860
gagctgcggg ctgagcaggg ccacccacag acctttgtgg acctgtgggc gctagtgtctg 1920
gagtcacatgc tgcacggcca ggtgatggat cagaagctgt caggacagag agccgccttg 1980
gaacgggggtc agaaccctct gcccctctac ttgagcctca atgtcaaaga gaacaatctg 2040
gagacactgg acttcaagga gtgggttgag ttctccccc atgaggtcgg ttctctgaag 2100
tacggggcct tegtccctcc tgagctcttc ggctccgagt tcttcatggg acggctgatg 2160
aggaggatcc cggagccccg gatctgcttt ctggaagcca tctggagcaa cattttctcc 2220
ctgaacctgc tggatgcctg gtatgacctc accagttctg gggagtcctg gaaacagcac 2280
atcaaggaca agaccaggag cttagagaag gagccccga ccacctcggg gacctcctcg 2340
cggctggagg cctcgtggct gcagccaggc acggcgctgg ccaggcatt taaaggcttc 2400
ctgacaggca ggcccccca ccagcgcagc cccaacttcc tccagggcct ccagctgcac 2460
caggactact gtagccacaa agacttctcc acctgggcag actaccagct tgactccatg 2520
cccagccagc tgacccccaa ggagccccgg ctctgcctgg tgagcgcgc ctacttcatc 2580
aacaccagct ctccctccat gttccggcca ggccgcaggc tggacctcat cctctccttc 2640
gactactccc tatctgcgcc cttcgaggca ctgcagcaga cggagctgta ctgccgggac 2700
cggggggctg ccttcccccg ggtggaaccc agccctcagg accagcacca gccaaaggaa 2760
tgccacctct tctcagaccc cgctgcccc gagggcccca tcctgctgca ctccccgtg 2820
gtcaatgcct ccttcaagga ccaactcagc cccggtgtcc agcgcagccc cgcagagctc 2880
caggggtggc aagtggatct caccggggcc acctgcccc acacctgtc caacatgacc 2940
tacaaggagg aagacttcca gcgcctgctg cggctcagtg actacaacgt gcagaccagc 3000
cagggtgcca tctgcaggc cctgaggacc gcgctgaagc accggactct agaggcgagg 3060
cctccaaggc cacagacctg aggttgctca gagctgcag gacctccag ggctgcggg 3120
cataacctga tctgtagctg ggctcagcca caggccttcc tggttggagt tctgggctct 3180
cccaggcctg ggtggcc 3197

```

&lt;210&gt; 15

&lt;211&gt; 2446

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2194717CB1

&lt;400&gt; 15

```

ggccggaacc cggcccgacc ccggcgcgcg cgcggcggag gacgaggaag agttgtggcg 60
aggcagatcc tgccccgtgg ccgcggccgt ctctgtaggtg tcccaccggc cttccgggtt 120
ccagcgccag gcctgggtgcc tgccccagga ggatgagcat ttgcattgcc tgcaccttat 180
cgtgcccttc cacctgctga agcagctgtg cctgccgctc ttgtgaactg cgaacttccc 240
cttacctct ctctctggct cgggagctgg ctccggccca ggaggtctcc gggagcgcct 300
gtcagtgcaa tgccgccccca tgtggtgtc acctccggc gcctgggctg cgccttggcc 360
tcctgcgggc tggcgccctgc gagacacaga ggaagtggc ttctgcacac agccccagtg 420
gcccgtctcg acaggagcgc ccggtgttc acccgtgcc tggccttgg ggacagaatc 480
gccctgggtg accagcacgg ccgccacacg tacaggggagc ttatttccc cagccttcgc 540
ctgtcccagg agatctgcag gctctgcggg tgtgtcggcg gggacctccg ggaggagagg 600
gtctccttcc tatgtgctaa cgacgcctcc tacgtcgtgg ccagtgggc ctcatggatg 660
agcggcggtg tggcagtacc cctctacagg aagcatccc cggcccagct ggagtatgtc 720
atctgcgact ccagagctc tgtggtcct gccagccagg agtacctgga gctcctgagc 780
ccggtgggtc ggaagctggg ggtcccgtg ctgccgctca caccagccat ctacactgga 840
gcagtagagg aaccggcaga ggtcccgtc ccagagcagg gatggaggaa caagggcgcc 900
atgatcatct acaccagtgg gaccacgggg aggcccaagg gcgtgctgag cacgcaccaa 960
aacatcaggg ctgtggtgac cgggctggc cacaagtgg catggacca agacgacgtg 1020
atcctccacg tgctcccgt gcaccacgtc catggtgtg tcaacgcgt gctctgtcct 1080
ctctgggtgg gagccacctg tgtgatgat cctgagttca gccctcagca ggtttgggaa 1140
aagttcttaa gttctgaaac gccgcggatc aatgtcttta tggcagtgcc tacaataatac 1200
accaagctga tggagtacta cgacaggcat tttaccagc cgcacgccc ggatttcttg 1260
cgtgcagttt gtgaagaaaa aattaggctg atggtctcag gctcagctgc cctgcccctc 1320
ccagtgtctg agaagtggaa gaacatcacg ggccacaccc tgctggagcg gtatggcatg 1380
accgagatcg gcatggctct gtcggggccc ctgaccactg ccatgcgcct gccaggttcc 1440
gtggggaccc cactgcctgg agtacagggt cgcatgtct cagaaaaccc acagagggaa 1500
gcctgtcct acaccatcca cgcagaggga gacgagagg ggaccaaggt gacccaggg 1560
tttgaagaaa aggaggggga gctgctggtg aggggaccct ccgtgttctg agaatactgg 1620
aataaaccag aagaaactaa gagtgcattc accctagatg gctggtttaa gacaggggac 1680

```

```

accgtggtgt ttaaggatgg ccagtactgg atccgaggcc ggacctcagt ggacatcatc 1740
aagactggag gctacaaggt cagcgccctg gaggtggagt ggcacctgct ggcccacccc 1800
agcatcacag atgtggctgt gattggagtt ccgatatga catggggcca ggggtcact 1860
gctgtggtga ccctccgaga aggacactca ctgtcccaca gggagctcaa agagtgggcc 1920
agaaatgtcc tggecccgtg cgcggtgccc tcggagctgg tgctggtgga ggagatcccg 1980
cggaaccaga tgggcaagat tgacaagaag gcgctcatca ggcacttcca cccctcatga 2040
cccggcagac tgggactgcg ggtctggtgg ggagcagcag acgtcccctt cacaccgaga 2100
accacggggg cccgtccaag acctggcctc ccttaaacct gaacccccca aatcaggtca 2160
cgtagaatca agaactgttt gggatgaaat caccatgtgg ggtccccagc ctggggccag 2220
ttgttgacgc tcaaggagac cgtccctggt gtcacctctg cctggtcacc gccgacctca 2280
tctgtgcagc cgggtgcagc cagccccctg cccacgtgc tgaggcacct cccgccccac 2340
agtgccttgc agttgccagg ctctccaggg caggtcccag aggtttccca caaaaaacaa 2400
ataaagactc cactggagga aacaagcccc tgtcaaaaaa aaaaaa 2446

```

&lt;210&gt; 16

&lt;211&gt; 2097

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473574CB1

&lt;400&gt; 16

```

atgctgtgtg ccctgctcct cctgcccagc ctcttggggg ccaccagggc cagccccacc 60
tcaggccccc aggagtgtgc aaagggtccc acggtgtggt gtcaggatct gcagacagct 120
gccaggtgcg gggctgtggg gtactgccaa gggcgcgat ggaacaaacc caccgcgaag 180
tctctgccct gcgacgtatg ccaggacata gcagccgccc ctggcaatgg gctgaacctt 240
gacgccacgg agtctgacat cctggctttg gtgatgaaga cctgtgagtg gctccccagc 300
caggagtctt cagccggatg caagtggatg gtggatgccc acagttcggc cctcctgagc 360
atgctccgtg gggccccgga cagtgtcccc gcacaggtgt gcacagcgt gcacctctgt 420
gagcgcgtgc agaggcacct ggccacctg aggcactct ccaaagagga cacctttgag 480
gctgtggctc cgctcatggc caatgggccc cttaccttcc acccccgcca ggccctgaa 540
ggagctctgt gccaagactg tgtacggcag cttgtggcca aaatcacccc agagaagggtg 600
tgcaagttca tccgtctgtg tggcaaccgg aggcgggccc gggcagtcga tgatgcctat 660
gccatcgtgc cgtccccaga gtgggacgcg gagaaccagg gcagcttctg caatgggtgc 720
aagaggctgc tcacggtgtc ctcccacaac ttggagagca agagcaccaa gcgagacatc 780
ctggtggcct tcaaggggtg ctgcagcatc ctgccgtgc cctatatgat ccagtgaag 840
cacttcgtca cccagtacga gcccggtgtc attgagagtc tcaaggacat gatggacccc 900
gtggctgtgt gcaagaagggt gggggcctgc cacggcccca ggacccccact gctgggcacc 960
gaccagtgtg ccctgggccc aagcttcttg tgcaggagcc aggaggccgc caagctgtgc 1020
aacgctgtgc aacactgcca gaagcatgta tggaagaga tgcacctcca cgctggggaa 1080
cacgcgtgac cgtggctgcc agagaccag agcctgctag cgaggcccat gaggtgggtg 1140
ctttcccat ccccatttca caaatgaaaa actgaagctc tgaggaggga ggctgggaag 1200
gagcagagct gaagttcaaa accaagtatt cctgatcccg aaagcctctc tcttaacaac 1260
ggtgccgcac agctttgccc ttgaaagcat ctctactgga ccggaacaca ctcatgtgcc 1320
ccgctccctg acccagccaa ggctgccctt tcacttccaa ggctgagatg ttgccggtgg 1380
tcccatgaga gctgcccac gggctcaggt gccctttac cttctgctgg atggacatct 1440
ggctgtgagc caggctgggg tcatggccgg ggtgagcgga ggcaggggct tggacggagg 1500
cttcgagggc ccatcactag taggttcatt acctcttgcc aacagccggg ggggtgggag 1560
ctgggtctcg ctcaggccag agcttctcaa cctggagtcc ctgggggtgg ctgcccgaag 1620
tgtgtatgac aagcacgtat ccctggacat ttccggggag aggtctgggg ctttggccac 1680
attctccaag ggctgtggtg cttcgagca gtcaccccc gatgtctcag ccactacagg 1740
gtccctctct ctccctgcac cccagacct ccgctgccct ggtaatgagc agaaggaaa 1800
tcttgggggt tgctcaaaag caggagagca aaatatgcca ggcaaaagct cccgggaaaa 1860
gccggaggag tctgggggtg ccaccgggat gtggagcagc gaggggcaag acggtgaaca 1920
cagccctcca gctgtctgag cctcagtttt ctaatctgta gaatggggat gatcatacct 1980
gcctcacaag aatgttgaga caattcacag agacgttctg gagccccctt ctcccagac 2040
cggcattcat gagtctgctg ggagcagaga acccatctca gagccccagc gggcacc 2097

```